



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTE DES MATHÉMATIQUES ET DE L'INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : GENIE LOGICIEL

Par :

BENCHOHRA Anouar

BOUKELIKHA Abdeldjalil

Sur le thème

Utilisation de l'Analyse Automatisée des Images pour la Reconnaissance des Expressions Faciales

Soutenu publiquement le .. / .. / 2020 à Tiaret devant le jury composé de :

M ^r MEZZOUG Karim	MAA	Université de Tiaret	Président
M ^r GAFOUR Yacine	MAA	Université de Tiaret	Encadreur
M ^r CHENINE Abdelkader	MAA	Université de Tiaret	Examineur

2019-2020

Dédicace

A tout respect et amour je dédie ce travail à
mes très chers parents, à mes frères et sœurs,
et tous mes amis.

Remerciement

Nous remercions en premier ALLAH qui nous a éclairé le chemin du savoir et qui nous a donné la volonté et la patience d'achever ce modeste travail de mémoire, notre grand salut sur le premier éducateur notre prophète Mohamed (satisfaction et salut de dieu soit sur lui).

Nous tiens à adresser mes remerciements aux nos parents qui a consenti des sacrifices et prodigué des encouragements tout au long de mes études.

*Nous adressons nos vifs remerciements et nos sincères gratitude, à notre Encadreur Monsieur **GAFOUR Yacine** qui nous a fait l'honneur d'avoir la charge d'encadrer notre travail de mémoire avec grande patience, pour la confiance qu'il a eu en notre projet et surtout pour ses orientations, ainsi que son aide précieuse et ses conseils pour réaliser ce mémoire.*

Nous Remercions également notre jury d'avoir accepté de juger notre travail et nous tenons à remercier également nos collègues pour leur aide de réalisation ce Modeste travail.

Abstract

The purpose of Facial Expressions recognition is to identify the emotions of a human being. The emotion can be captured either from the face or from verbal communication. In this work, we focus on the identification of human emotions from facial expressions. The recognition of facial emotions is one of the useful tasks and can serve as a basis for many real-time applications. It can be used in many interesting applications such as security monitoring, patient treatment in the medical field, marketing research, e-learning, etc.

The facial expression recognition system is implemented using the Wide Margin Separator (SVM) using the Local Binary Pattern (LBP) descriptor. The facial expression recognition system is evaluated on the CK+ database, which contains seven categories of facial expressions: "happy", "sad", "surprise", "contempt", "fear", "disgust", "anger". The recognition results of our system are very astonishing

Table des matières

1	Liste des figures	I
	Liste des tableaux	I
1.1	.Introduction.....	2
1.2	Qu'est-ce que la vision par ordinateur ?.....	2
1.3	.Complexité du système de la vision par ordinateur	2
1.4	Système visuel humain	3
1.5	Comment les humains comprennent-ils le contenu de l'image ?.....	3
1.6	Pourquoi est-il difficile pour les machines de comprendre le contenu des images ? 3	
1.7	.Conclusion	4
1	. Introduction	5
2	Propriétés des caractéristiques de l'image	5
3	Détection, description des caractéristiques de l'image	5
3.1	Extraction des points clés.....	5
3.1.1	Détecteurs de caractéristique invariants à la rotation	6
3.1.1.1	Détecteur Hessian	6
3.1.1.2	Détecteur Harris	6
3.1.1.3	Laplacian of Gaussian (LoG).....	8
3.1.1.4	Difference of Gaussian (DoG).....	8
3.2	Les descripteurs	9
3.2.1	SIFT	9
3.2.2	SURF	10
3.2.2.1	Détection et localisation de points clés.....	10
3.2.2.2	Orientation assignement.....	10
3.2.3	HOG.....	11

3.2.4	BinaryRobust Independent Elementary Features	12
3.2.5	Local Binary Pattern	13
4	.Apprentissage automatique pour la reconnaissance des émotions	17
5	Machine Learning pour la classification des images.....	18
5.1	Apprentissage supervisé.....	18
5.1.1	Classification	18
5.1.1.1	Support Vector Machines	18
5.1.1.2	K-NearestNeighbor	22
5.1.1.3	Naive Bayes	23
5.2	Apprentissage non supervisé.....	23
5.2.1	Clustering.....	23
5.2.1.1	K-means	23
5.2.1.2	Ascending hierarchical classification	25
6	Réduction de dimension en utilisant ACP.....	26
7	Les réseaux de neurones et Deeplearning	26
7.1	Classification des réseaux de neurones.....	26
7.1.1	Types de réseau de neurones profond.....	28
7.1.1.1	Convolutional neural networks	28
7.1.2	Différence entre l'apprentissage automatique et l'apprentissage profond.....	30
8	Quand utiliser ML ou DL?	31
9	Conclusion.....	32
1	1. Introduction	33
2	Reconnaissance des émotions ?.....	33
3	Approches de la reconnaissance des émotions.....	34
3.1	Approche proposée	35

1.1	Extraction de caractéristiques.....	35
1.1.1	Descripteur LBP et Processus du système de la reconnaissance des émotions	35
3.1.1	Expériences et résultats	42
3.1.1.1	Basse de donnée.....	42
2.2.2	Résultats obtenus	42
4.	Conclusion	50
	Conclusion générale	51
	Bibliographie	52
	Web bibliographie	52

1 Liste des figures

Figure 2.1 :Détection des coins en fonction des valeurs propres	7
Figure 2.2: Fenêtre de détection avec cellule et bloc	11
Figure 2.3: représentation graphique de la méthode LBP	13
Figure 2.4: matrice binaire pour la méthode LBP	15
Figure 2.5: processus d'opération LBP dans l'entrée et la sortie	15
Figure 2.6: exemples de voisinage utilisés pour définir une texture et calculer un motif binaire local (LBP) Représentation des caractéristiques de l'image.....	15
Figure 2.7: Méthode SURF pour détecter les points caractéristiques	16
Figure 2.8: classification a multi class.....	19
Figure 2.9: Hyperplan à marge maximale et marges pour un SVM formé avec des échantillons de deux classes. Les échantillons sur la marge sont appelés les vecteurs de support.....	20
Figure 2.10: Hyperplan pour optimisation de la marge maximale et marges pour un SVM formé avec des échantillons de deux classes. Les échantillons sur la marge sont appelés les vecteurs de support.....	20
Figure 2.11: classification des données basées sur la méthode KNN	22
Figure 2.12/ Algorithme de K-means. Les exemples d'entraînement sont représentés sous forme de points et les centres de gravité de cluster sont représentés par des croix.....	24
Figure 2.13: représentation graphique du clustering Hierarchical CHA.....	25
Figure 2.14: Comparaison des produits biologiques et du réseau neuronal	27
Figure 2.15: L'architecture d'un réseau neuronal, chaque couche est constituée de nœuds. Le nombre de couches masquées est facultatif.	28
Figure 2.16: modèle pour une architecture réseau.....	30
Figure 3.1: Architecture d'un système de reconnaissance des expressions faciales	36
Figure 3.2: Exemple sur la construction les 4 niveaux de coté droite.....	37
Figure 3.3: Modèle de découpage d'image	38
Figure 3.4: Concaténation des régions pour avoir le descripteur	39
Figure 3.5: Les histogrammes sont combinés à tous les niveaux à un seul angle D et une SVM est utilisée pour la classification.	40
Figure 3.6: la classification image a travers PCA	41
Figure 3.7: Représentation du temps de modèle apprentissage aux 3 niveaux sans PCA... 44	44

Figure 3.8: Représentation du précision de modèle apprentissage aux 3 niveaux sans PCA	45
Figure 3.9: Représentation du temps de modèle apprentissage aux 3 niveaux avec PCA ..	46
Figure 3.10: Représentation du précision de modèle apprentissage aux 3 niveaux avec PCA	47
Figure 3.11: Interface de l'application de reconnaissance des expressions faciales	50

Liste des tableaux

Tableau 3.1: tableau représente la base de données CK+ utilisé.....	42
Tableau 3.2: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 0.....	43
Tableau 3.3: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 1.....	43
Tableau 3.4: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 2.....	44
Tableau 3.5: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 3.....	44
Tableau 3.6: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 1.....	45
Tableau 3.7: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 2.....	45
Tableau 3.8: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 3.....	46
Tableau 3.9: Matrice de précision utilisant le descripteur SPM et le classifieur SVM avec et sans PCA au niveau 1.....	48

Liste des abbreviations:

HOG: Histogram of oriented gradients

KNN: K-nearest neighbors.

LBP : Local binary pattern.

SVM : Support vector machine.

ACP :Analyse en composantes Principales

CK+ :Extended Cohn Kanade

CNN :Convolutional Neural Networks

HMM :Hidden Markov Model

SIFT :Scale-Invariant Feature Transform

SURF :speed up robust features

BRISK :binary robust invariant scalable keypoints

NB :Naive Bayes

ML :machine learning

DP :deep learning

Introduction générale

La reconnaissance de l'expression faciale est un concept intéressant et qui a d'importantes applications dans de nombreux domaines tels que (Sécurité, Interaction Homme-Machine, Marketing, etc.), grâce a, la détection de différentes possibilités d'expressions (bonheur, tristesse, surprise, colère, peur et dégoût). Les travaux de ce mémoire visent à concevoir un système de reconnaissance des expressions faciales. Le système de reconnaissance des expressions faciales est caractérisé par trois étapes nécessaires. Les étapes sont l'enregistrement du visage, l'extraction des caractéristiques et classification.

L'objectif de ce projet de fin d'étude est de concevoir et de réaliser une application de reconnaître des expressions faciales en utilisant le descripteur *Local Binary Pattern* (LBP) .Ce mémoire est constitué de quatre chapitres, et organisé comme suit :

Dans le premier chapitre, nous présentons de façons générales la vision par ordinateur en présentant la complexité et la difficulté des machines pour comprendre le contenu des images.

Le second chapitre est consacré à l'étude des différentes étapes du système de reconnaissance faciale. Dans ce chapitre nous présentons l'état de l'art sur l'extraction des caractéristiques ainsi que la machine learning.

Le dernier chapitre porte sur l'implémentation et conception de notre système de reconnaissance, où nous exploitant la base de données CK+ pour évaluer notre système proposé. À la fin de ce chapitre, nous présentons les résultats obtenus et discutés. Nous terminerons enfin par une conclusion générale dans laquelle nous résumerons le travail réalisé.

I. Introduction à la vision par ordinateur

1.1 .Introduction

La vision de l'ordinateur est une discipline informatique qui cherche à développer des techniques pour aider les ordinateurs à "voir" et à comprendre le contenu des images numériques telles que les photographies et les vidéos. L'objectif de ce chapitre est d'expliquer la vision par ordinateur, et comment prendre des décisions utiles des objets physiques réels.

1.2 Qu'est-ce que la vision par ordinateur ?

La vision par ordinateur (CV) est un domaine interdisciplinaire qui permet aux ordinateurs de comprendre, de traiter et d'analyser des images. Elle utilise un algorithme qui peut traiter des images et vidéo. Cela aide l'ordinateur à afficher, interpréter et traiter les images sous la forme de la vision humaine et à produire ensuite un résultat acceptable (Radu Horaud, Olivier Monga, 01/06/1995)

1.3 .Complexité du système de la vision par ordinateur

L'une des raisons est que nous n'avons pas une bonne compréhension du fonctionnement de la vision humaine. L'étude de la vision biologique nécessite une compréhension des organes de perception comme les yeux, ainsi que l'interprétation de la perception dans le cerveau. De nombreux progrès ont été réalisés, tant dans la cartographie du processus que dans la découverte des astuces et des raccourcis utilisés par le système, même si, comme toute étude impliquant le cerveau, il reste un long chemin à parcourir.

Une autre raison pour laquelle il s'agit d'un problème si difficile est la complexité inhérente au monde visuel. Un objet donné peut être vu de n'importe quelle orientation, dans toutes les conditions d'éclairage, avec n'importe quel type d'occlusion à partir d'autres objets, et ainsi de suite. Un vrai système de vision doit être capable de «voir» dans n'importe laquelle d'un nombre infini de scènes et d'extraire encore quelque chose de significatif. Les

ordinateurs fonctionnent bien pour des problèmes étroitement limités, pas pour des problèmes sans limites comme la perception visuelle.

1.4 Système visuel humain

Le système visuel crée une image mentale de l'environnement et cela contribue à notre capacité à nous déplacer efficacement dans l'espace physique et à communiquer avec des gens et recueillent des informations sensorielles qui nous aident à comprendre le monde qui nous entoure.

1.5 Comment les humains comprennent-ils le contenu de l'image ?

La rétine de l'œil, qui contient 150 millions de cellules en bâtonnets et cônes sensibles à la lumière, est en fait une excroissance du cerveau. Dans le cerveau lui-même, les neurones consacrés au traitement visuel se comptent par centaines de millions et occupent environ 30% du cortex, contre 8% pour le toucher et seulement 3% pour l'audition. Chacun des deux nerfs optiques, qui transportent les signaux de la rétine au cerveau, se compose d'un million de fibres, chaque nerf auditif en porte à peine 30 000. Le cortex visuel primaire coopère avec d'autres parties du cerveau pour traiter différents composants d'informations visuelles telles que l'orientation verticale ou horizontale, la couleur, la taille, la forme, le mouvement et les distinctions entre les objets qui se chevauchent

La capacité d'identifier les images vues si brièvement peut aider le cerveau à décider où focaliser les yeux, qui s'élancent d'un point à l'autre en de brefs mouvements appelés fixations environ trois fois par seconde. Décider où déplacer les yeux peut prendre de 100 à 140 millisecondes, donc une compréhension très rapide doit avoir lieu avant cela

1.6 Pourquoi est-il difficile pour les machines de comprendre le contenu des images ?

Nous pouvons rencontrer de nombreux objets chaque jour et sans effort, nous pouvons les reconnaître presque immédiatement. Par exemple, lorsque vous voyez une table, vous n'attendez pas quelques minutes avant de vous rendre compte que c'est bien une table. D'un autre côté, les ordinateurs trouvent qu'il est très difficile d'accomplir cette tâche. Les

chercheurs travaillent depuis des décennies pour comprendre pourquoi les ordinateurs ne sont pas aussi performants que nous ?

Pour obtenir une réponse à cette question, nous devons comprendre comment les humains le font. Nous avons parlé dans la section précédente comment les données visuelles pénètrent dans le système visuel humain et comment notre système les traite. Le problème resté toujours est que nous ne comprenons pas complètement comment notre cerveau analyse, organise et reconnaît ces données visuelles. Par ailleurs, la reconnaissance des objets par l'ordinateur commence par l'extraction de certaines caractéristiques des images. Ensuite, l'ordinateur va exploiter ces caractéristiques à l'aide des algorithmes d'apprentissage automatique.

1.7 .Conclusion

Nous avons présenté dans ce chapitre le concept général de la vision par ordinateur, en montrant son importance majeure. Nous avons également mis le point sur le concept de système visuel humain et la machine et la difficulté de la dernière à comprendre le contenu de l'image. Dans le chapitre qui suit, nous abordons le traitement des images et la reconnaissance de formes ainsi que la détection et description des caractéristiques des images.

II. Traitement des images et reconnaissance de formes

1 . Introduction

Dans ce chapitre, nous abordons le domaine de traitement des images. Ce domaine est la procédure qui peut être effectuée sur les images. Avant de traiter les images, il faut exploiter les images d'entrées. C'est une l'étape d'acquisition qui consiste à convertir les vues réelles en d'autres images numériques.

2 Propriétés des caractéristiques de l'image

Image numérique se compose de pixels : Un pixel est le très petit élément d'une image. Le pixel représente un point de l'image (dans un espace colorimétrique prédéfini), Par exemple, dans une image en noir et blanc (N&B), un pixel sera représenté par un bit (0 = noir, 1 = blanc). Une image en niveau de gris est codée sur 8 bits, ce qui peut donner 256 valeurs en niveaux de gris.

Une image couleur est représentée par exemple dans l'espace RGB (256 nuances de rouge, 256 nuances de vert, 256 nuances de bleu). Chaque pixel est représenté par 3 octets = 3 x 8 bits qui peut générer 16,8 millions de couleurs.

3 Détection, description des caractéristiques de l'image

3.1 Extraction des points clés

Les points d'intérêt sont également appelés entités ponctuelles, font référence aux pixels ayant des propriétés uniques dans une image. Ces point sont des caractéristiques essentielles d'une image. Ils présentent les avantages de l'invariance de rotation et ne changent pas dans des conditions de lumière. Le détecteur de points d'intérêt est une technique ou un système permettant d'identifier les positions d'une image qui peut être reproduit de manière stable dans d'autres images, et est couramment utilisé dans le domaine de la vision par ordinateur.

3.1.1 Détecteurs de caractéristique invariants à la rotation

La plupart des algorithmes de correspondance d'images et de reconnaissance d'objets doivent gérer le problème de la rotation de l'image.. Une façon de résoudre ce problème est de construire des vecteur qui contient des points clés invariants à la rotation.

3.1.1.1 Détecteur Hessian

La détection des points clés en Hessian est l'un des détecteur le plus célèbre où les points détectés sont les coordonnées des pixels de l'image qui présentent des valeurs les plus extrêmes. En considérant le changement continu dans l'espace tridimensionnel des coordonnées $p = (x, y)$ et la valeur du pixel $I(p)$, l'image peut être vue comme une surface courbe. L'existence ou non d'une valeur extrême à la coordonnée p de la surface quadrangulaire est exprimée dans la matrice Hessian H présentée ci-dessous

$$H(x, y) = \begin{pmatrix} I_{XX}(x, y) & I_{XY}(x, y) \\ I_{XY}(x, y) & I_{YY}(x, y) \end{pmatrix} \quad (2.1)$$

σ est la taille du noyau du filtre gaussien

$I_{XY}(x, y, \sigma)$ est la dérivée d'image du second ordre (M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, 2016).

$$\det(H) = I_{XX}I_{YY} - I_{XY}^2 \quad (2.2)$$

La valeur de $\det(H)$ est utilisée comme point clé si le score atteint le seuil (*threshold*). Le point clé ayant le score le plus élevé est laissé dans une zone limitée comme 3 ou 3 pixels pour les points clés détectés, et un traitement de suppression non maximale est effectué pour éliminer les autres points clés voisins ayant des scores non maximaux.

3.1.1.2 Détecteur Harris

L'algorithme de détecteur Harris d'utilise une fenêtre locale glissante afin de décider si le niveau de gris change de manière significative. Si la valeur de gris dans la fenêtre a un décalage important, alors il y a des points d'angle dans la zone de la fenêtre matrice suivant représente L'algorithme.

$$M = \sum_{u,v} w(u,v) * \begin{bmatrix} I_x^2(x,y) & I_x I_y(x,y) \\ I_x I_y(x,y) & I_y^2(x,y) \end{bmatrix} \quad (2.3)$$

Les étapes de harris detecteur :

1. calculer la Dérivées
2. calculer la Dérivées au carré
3. calculer la Dérivées au carré, filtrées avec gaussienne
4. Calculer fonction des valeurs propres de M

Détecteur de Harris: interprétation :

Classification des points en fonction des valeurs propres

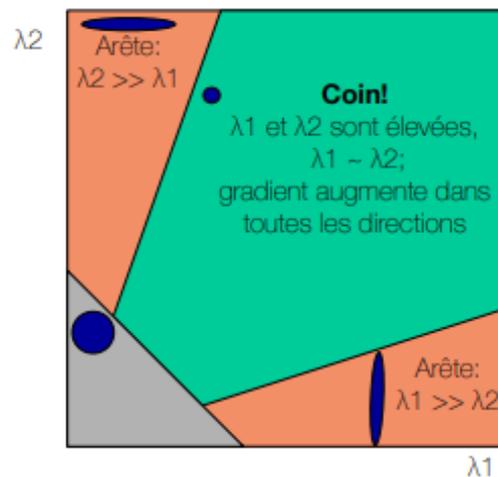


Figure 2.1 : Détection des coins en fonction des valeurs propres

Région uniforme: λ_1 et λ_2 sont petites: gradient constant dans toutes les directions

Détecteur de Harris: math

En pratique, nous n'avons pas besoin de calculer les valeurs propres

$$R = \frac{\det M}{\text{tr } M} \quad (2.4)$$

$$\det M = \lambda_1 \lambda_2 \quad (2.5)$$

$$tr = \lambda_1 + \lambda_2 \quad (2.6)$$

$$\det M = m_{11}m_{22} - m_{21}m_{12} \quad (2.7)$$

$$tr = m_{11} + m_{22} \quad (2.8)$$

De cette façon, l'algorithme de détection des coins de Harris peut être divisé en trois étapes :

- Calculer R pour tous les points dans l'image
- Appliquer: $R > \text{seuil}$
- Retenir les maximums locaux seulement

3.1.1.3 Laplacian of Gaussian (LoG)

Laplacian of Gaussian est une combinaison de filtre gaussien et de filtre laplacien. Après avoir lissé l'image avec un filtre gaussien pour réduire le bruit, le filtre laplacien extrait le contour. Puisque le filtre laplacien agit comme une dérivée seconde. Il a la caractéristique que le bruit est facilement accentué. Ceci est supprimé en lissant l'image au préalable avec un filtre gaussien.

Formule de calcul du filtre LoG :

$$LoG(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2.9)$$

Cette équation peut être obtenue par différenciation du second ordre de l'équation du filtre gaussien . (R. Fisher, S. Perkins, A. Walker and E. Wolfart., 2003)

3.1.1.4 Difference of Gaussian (DoG)

Difference of Gaussian est une opération où la différence pixel par pixel de l'image en échelle de gris convoluée gaussienne est obtenue en premier, l'image en échelle de gris est étouffée par convolution avec le noyau gaussien avec une certaine déviation standard σ

Le filtre générant cette différence d'image est appelé filtre DoG. Les filtres DoG sont identiques pour la procédure de filtre LoG, Il est également utilisé à la place du filtre LoG, car il nécessite un petit calcul.

La formule du filtre DoGest:

$$DoG(x, y) = G(x, y, \sigma_1) - G(x, y, \sigma_2) \quad (2.10)$$

Où se trouvent deux images gaussiennes avec des valeurs

$$G(x, y, \sigma_1), G(x, y, \sigma_2) \quad (2.11)$$

d'écart type différentes. (M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, 23 February 2016)

3.2 Les descripteurs

3.2.1 SIFT

SIFT (*Scale-Invariant Feature Transform*) est un algorithme pour l'identification et la définition des points caractéristiques. Un avantage majeur de ce descripteur est la possibilité de définir des quantités de caractéristiques robustes contre la rotation, la taille, les changements d'éclairage, etc. Il est utilisé pour la reconnaissance et la détection d'objets, ainsi que pour le développement de photographies panoramiques. La procédure de traitement SIFT est résumée approximativement comme suit :

1/ Détection des points caractéristiques :

-Déterminez les échelles et les points caractéristiques par traitement DoG.

-Les points caractéristiques sont réduits et l'estimation de la position des sous-pixels est effectuée.

2/ Description des caractéristiques :

- Calculez la direction du gradient afin d'obtenir l'orientation.

- Calculez l'histogramme de direction du gradient et décrivez la quantité d'entités.

(Alshazly, M. Hassaballah Aly Amin Abdelmgeid Hammam A., 2016)

3.2.2 SURF

SURF(*Speed Up Robust Features*) est un algorithme d'extraction de points caractéristiques proposé par Bay. Les étapes de l'algorithme SURF sont similaires à l'algorithme SIFT, mais sa vitesse est plusieurs fois supérieure à celle de l'algorithme SIFT. SURF est basé sur le détecteur Hessian. (Herbert Bay, Andreas Ess, Tinne Tuytelaars, 2008)

Les étapes de SURF sont :

3.2.2.1 Détection et localisation de points clés

Dans un premier temps, un détecteur des points est basé sur la matrice de Hesse. Il est utilisé pour localiser les points clés. Le déterminant de la matrice de Hesse est utilisé pour sélectionner l'emplacement et l'échelle des points clés. Plus précisément, pour un point donné $p(x, y)$ sur l'image I , $M(i, j)$ représente une matrice de Hesse d'échelle σ au point clé p , qui se définit comme suit:

$$M(i, j) = \begin{pmatrix} \mu_{11}(i, j) & \mu_{12}(i, j) \\ \mu_{21}(i, j) & \mu_{22}(i, j) \end{pmatrix} \quad (2.12)$$

Parmi eux, $\mu_{11}(i, j)$ est la convolution de la dérivée gaussienne du second ordre de l'image I au point clé. Cependant, SURF utilise la dérivée gaussienne approximative du second ordre au lieu d'utiliser un filtre gaussien.

3.2.2.2 Orientation assignement

Afin d'obtenir l'invariance de la rotation d'image, chaque point clé se voit attribuer une direction constante basée sur ses propriétés d'image locales. Le descripteur de point-clé peut alors être exprimé par rapport à cette direction. L'algorithme utilisé pour trouver la direction comprend les étapes suivantes:

- 1) L'échelle du point clé est utilisée pour sélectionner l'image floue gaussienne avec l'échelle la plus proche.
- 2) Calculez la taille et la direction du gradient pour chaque pixel de l'image à cette échelle.

3) L'histogramme est construit à partir de la direction du gradient des pixels dans la zone locale autour du point clé, couvrant une plage de direction de 360.

4) Le pic le plus élevé de l'histogramme d'orientation locale correspond à la direction dominante du gradient local. En outre, tous les autres pics locaux à moins de 80% du pic le plus élevé sont également considérés comme des points clés dans cette direction.

3.2.3 HOG

HoG (histogrammes de dégradés orientés) est un histogramme qui représente la direction du gradient d'un ensemble de pixel dans un bloc locale (cellule). Cet histogramme représente un descripteur de caractéristiques utilisé pour détecter des objets en vision par ordinateur et en traitement d'image. (NAZIR, Muhammad, JAN, Zahoor, et SAJJAD, Muhammad, 2018)HOG se caractérise par sa robustesse à l'échelle de l'image. Son algorithme est :

1/ Divisez l'image d'entrée en plusieurs blocs. Ils divisent également chaque bloc en cellules.

Lors de la division une image en des blocs.

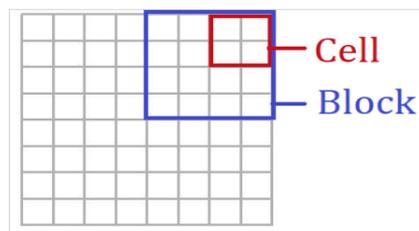


Figure 2.2: Fenêtre de détection avec cellule et bloc

2/ Différencier l'image d'entrée.

$$I_x(x, y) = I(x + 1, y) - I(x, y) \quad (2.13)$$

$$I_y(x, y) = I(x, y + 1) - I(x, y) \quad (2.14)$$

3/ Obtenir la force du gradient et la direction du gradient à partir de l'image différentielle

$$|I| = \sqrt{I_x^2 + I_y^2} \quad (2.15)$$

$$\theta = \tan^{-1} \frac{I_y}{I_x} = \text{atan2}(I_y, I_x) \quad (2.16)$$

4/ Quantifiez la direction du gradient dans 9 directions (20 degrés de 0 à 180 degrés).

5/ Calculez l'histogramme de direction du dégradé en pondérant chaque cellule avec intensité.

6/ Normaliser chaque bloc et calculer le montant de la fonction.

$$h(n) = \frac{h(n)}{H} \quad (2.17)$$

H(n) Est le nième histogramme de direction du gradient.

Le dénominateur H est la somme des caractéristiques. HOG dans un bloc et est calculé par la formule suivante.

$$H = \sqrt{(\sum_{k=1}^{m \times m \times N} h(k)^2) + \epsilon} \quad (2.18)$$

M est la taille de la cellule et N est le nombre de directions de dégradé $\epsilon = 1$

6/ Achèvement des caractérisations HoG en combinant tous les histogrammes.

3.2.4 BinaryRobust Independent Elementary Features

BRIEF (*BinaryRobust Independent Elementary Features*) est un méthode simple de définir des caractéristiques binaires à partir du signe de la différence de luminosité de paires de pixels sélectionnées de façon aléatoire à l'intérieur de l'image de patch autour du point clé. La paire de pixels I choisie au hasard à l'intérieur de l'image du patch. (Michael Calonder, Vincent Lepetit, Mustafa Özuysal, 2012)

$I(P_i)$, $I(P_j)$ sont les valeurs des pixels P_i , P_j , alors que, la fonction binaire τ pour le bit I peut être obtenue par

$$\tau = \begin{cases} 1, & \text{si } I(P_j) < I(P_i) \\ 0, & \text{autrement} \end{cases} \quad (2.19)$$

En bref, L'image du patch est lissée à l'avance avec un filtre gaussien pour minimiser les effets du bruit. Un traitement à grande vitesse est donc possible, car le résultat de la fonction peut être représenté simplement en mesurant la différence de luminosité des paires de pixels dans l'image de patch pour le nombre de bits (nombre de dimensions).

3.2.5 Local Binary Pattern

LBP (*Local Binary Patterns*) La technique est proposée par (Ojala, T, Pietikäinen, & Harwood, D, 1996). Le principal concept derrière Local Binary Pattern (LBP) est d'utiliser des codes binaires pour définir le voisinage des éléments d'image. En général, est un algorithme utilise une approche pour rechercher ses attributs locaux et pour définir les caractéristiques de chaque partie de l'image.

L'algorithme combine les méthodes statistiques et les méthodes structurelles. Cet algorithme est considéré comme un système facile et rapide, et a donné d'excellents résultats dans de nombreuses études.

LBP, comme son nom l'indique, est une caractéristique de la représentation locale d'une image. En comparant chaque pixel avec ses voisins, il est constitué de valeurs relatives.

Les principales caractéristiques du LBP sont les suivantes :

- 1.- Faible coût de calcul
- 2- La valeur de gris de l'image résiste aux fluctuations

* Étapes de la méthode LBP *

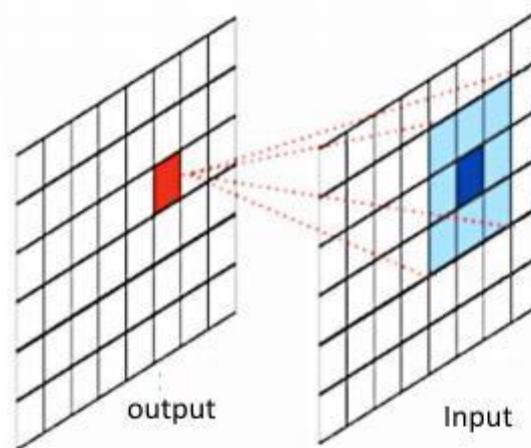


Figure 2.3:représentation graphique de la méthode LBP

- 1- Convertir l'image en gris spatial.
- 2- Sélectionnez des quartiers P autour du pixel central pour chaque pixel (gp) de l'image.

3- Prenez un pixel central (gc) et définissez-le comme le seuil P de son voisin.

4- Si la valeur du pixel adjacent est supérieure ou égale à la valeur du pixel central, définissez cette valeur à 1, sinon mettez la 0.

5- Calculez maintenant la valeur de LBP : écrivez un nombre binaire dans le sens inverse des aiguilles d'une montre, composé de nombres adjacents au pixel central. Ce nombre binaire (ou son équivalent décimal) est appelé le code du pixel central LBP, et est également utilisé comme texture locale pour sélectionner des caractéristiques.

$$LBP(gp_x, gp_y) = \sum_{p=0}^{p=7} S(gp - gc) \times 2^p \quad (2.20)$$

> Formule LBP uniforme

gc- la valeur d'intensité du pixel central

gp- l'intensité du pixel voisin d'indice p

La fonction S peut être exprimée comme:

$$S(x) = \begin{cases} 1, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases} \quad (2.21)$$

P est nombre de points d'échantillonnage sur un cercle de rayon R (voisinage circulaire).

Quantification des méthodes de contrôle P.

R- détermine la résolution spatiale de la méthode ou de l'opérateur.

Exemple

La formule suivante peut être utilisée pour décrire la taille de cette fenêtre (3x3) : le rayon du cercle est égal à $(2 * R + 1)$, si le rayon est de 1, vous obtiendrez une matrice de 3x3.

Prenons l'exemple de 8 échantillons voisins (P = 8). Les coordonnées des différents points d'échantillonnage peuvent être exprimées comme suit :

P = 8

p = 0, 1, 2 ..., P-1

$$gc_x = gc_y = 1$$

0	1	1
0		1
1	1	1

Figure 2.4: matrice binaire pour la méthode LBP

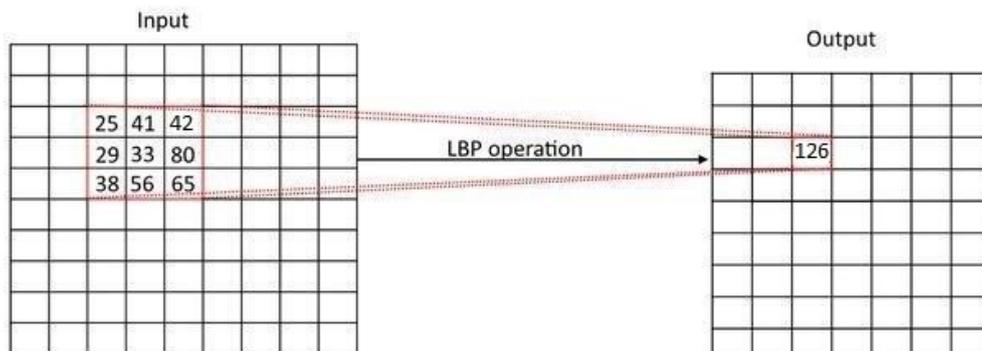


Figure 2.5: processus d'opération LBP dans l'entrée et la sortie

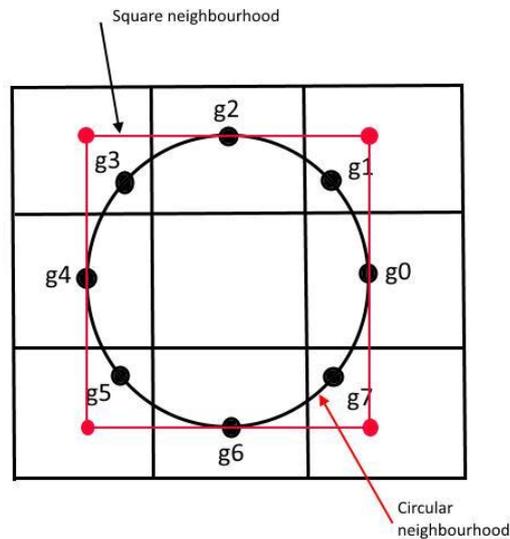


Figure 2.6: exemples de voisinage utilisés pour définir une texture et calculer un motif binaire local (LBP) Représentation des caractéristiques de l'image

Le modèle du sac de mots (modèle BoW) (Ioannis Pratikakis, Anastasia Bolvinou, Bassilios Gatos, and StavrosPerantonis, 2011) est apparu pour la première fois dans le traitement du langage naturel (*Natural Language Processing*) et la recherche d'informations.

Le modèle ignore la grammaire et l'ordre des mots du texte et le traite comme une collection de vocabulaires multiples, et chaque mot apparaît séparément dans le document.

le modèle du sac de mots et proposer pour identifier les images (Sac de points clés). Les mots dans l'image sont définis comme le vecteur caractéristique d'une zone d'image, et le modèle BoW de l'image est "l'histogramme obtenu à partir de tous les vecteurs caractéristiques de la zone d'image dans l'image".

Les étapes suivantes concernent le modèle BoW utilisé pour la classification des images afin d'extraire les histogrammes des caractéristiques :

- Extraction de caractéristiques

Extraire le vecteur de caractéristique dans le bloc d'image de l'échantillon d'entraînement, la méthode d'extraction du vecteur de caractéristique peut être SIFT, SURF etc. La fonction d'image descriptive générée à cette étape doit être invariante aux changements d'éclairage, aux déformations, aux collisions, etc.

La figure suivant montre une image des points de caractère extraits à l'aide de la méthode SURF



Figure 2.7: Méthode SURF pour détecter les points caractéristiques

- Construire un dictionnaire (Vocabulaire)

Grâce à l'extraction de la fonction dans la phase précédente, nous avons obtenu les caractéristiques de toutes les images de l'exemple de formation. Par exemple, s'il y a N images de formation, nous pouvons éventuellement obtenir un ensemble de N points de caractéristiques en utilisant la méthode SURF pour extraire les caractéristiques de l'image. Ensuite, nous allons construire un dictionnaire en utilisant les points caractéristiques. Le processus de construction d'un dictionnaire implique certains algorithmes de regroupement, généralement la méthode *k-means*. On choisit d'abord n centres de grappes, et on utilise un algorithme itératif pour déterminer les n derniers centres de grappes.

- Calculer l'histogramme des caractéristiques de l'image

N nouveaux centres de grappes sont obtenus après avoir passé l'algorithme de mise en grappes K-means, c'est-à-dire qu'une base est obtenue pour l'histogramme caractéristique

Ensuite, nous regroupons les points caractéristiques de l'image dans n dictionnaires (c'est-à-dire n centres de groupes) qui sont déjà formés, et nous comptons le nombre de points caractéristiques qui vont dans chaque dictionnaire. Nous pouvons ainsi obtenir un histogramme des caractéristiques d'une image.

4 .Apprentissage automatique pour la reconnaissance des émotions

La détection des émotions humaines est nécessaire à de nombreuses fins, afin que les systèmes artificiels intelligents modernes puissent imiter et mesurer les réactions du visage. Cela peut être utile pour prendre des décisions éclairées sur la reconnaissance des objectifs, la promotion d'offres ou les menaces à la sécurité. Reconnaître des émotions à partir de photographies ou de vidéos est une activité simple pour l'œil humain, mais elle s'avère très difficile pour les machines qui nécessite de nombreuses techniques de traitement d'images pour l'extraction d'informations.

Toute détection ou reconnaissance de l'apprentissage d'une machine implique la formation d'algorithmes, puis leur test sur un ensemble de données acceptable. Ce chapitre traite de quelques algorithmes d'apprentissage machine ainsi que des techniques d'extraction de caractéristiques qui nous aideraient à reconnaître les émotions humaines avec précision

5 Machine Learning pour la classification des images

5.1 Apprentissage supervisé

L'apprentissage supervisé est une approche d'apprentissage automatique qui nous permet d'apprendre à partir des données d'entraînement ou de construire un modèle d'apprentissage basé sur cet ensemble de données, et d'en déduire de nouveaux exemples. Les données d'apprentissage sont composées d'objets en entrée (généralement des vecteurs) et de la sortie prévue. La sortie de la fonction peut être une valeur continue (appelée analyse de régression), ou une marque de classification (appelée classification) peut être attendue.

5.1.1 Classification

Une autre forme courante d'apprentissage supervisé est la classification (également appelée classification, est une classification automatique de données inconnues

La classification est également couramment utilisée dans l'intelligence artificielle, ainsi que la classification des images qui différencient les artefacts tels que la reconnaissance des images et les émotions

La classification est une activité standard de l'intelligence artificielle qui s'exerce depuis des décennies. Il existe plusieurs algorithmes d'apprentissage machine qui effectuent la classification à cette fin, et divers tels que l'arbre de décision qui utilise la classification basée sur des règles de décision et l'approche K plus proche voisin et Support Vector Machines etc. Comme le rôle d'identification et de classification des points caractéristiques est également un genre approprié pour l'apprentissage profond, l'apprentissage profond est utilisé pour classifier des éléments ambigus qui ne peuvent pas être déterminés facilement.

5.1.1.1 Support Vector Machines

Le SVM est un algorithme supervisé dans l'apprentissage machine. Il peut essentiellement être utilisé pour des tâches de classification et de régression, mais en pratique, il est surtout utilisé pour des tâches de classification.

Le SVM multiclasse vise à attribuer des étiquettes aux instances en utilisant des machines à vecteurs de support, où les étiquettes sont tirées d'un ensemble fini de plusieurs éléments.

L'approche dominante pour ce faire est de réduire le problème multiclass unique en plusieurs problèmes de classification binaire. Les méthodes courantes pour une telle réduction comprennent :

Construire des classificateurs binaires qui distinguent entre l'un des labels et le reste (un contre tous) ou entre chaque paire de classes (un contre un). La classification de nouvelles instances pour le cas un contre tous se fait par une stratégie "winner-takes-all", dans laquelle le classificateur ayant la fonction de sortie la plus élevée attribue la classe (il est important que les fonctions de sortie soient calibrées pour produire des scores comparables). Pour l'approche un contre un, la classification est effectuée par une stratégie de vote à gains maximums, dans laquelle chaque classificateur assigne l'instance à l'une des deux classes, puis le vote pour la classe assignée est augmenté d'une voix, et enfin la classe avec le plus de voix détermine la classification de l'instance.

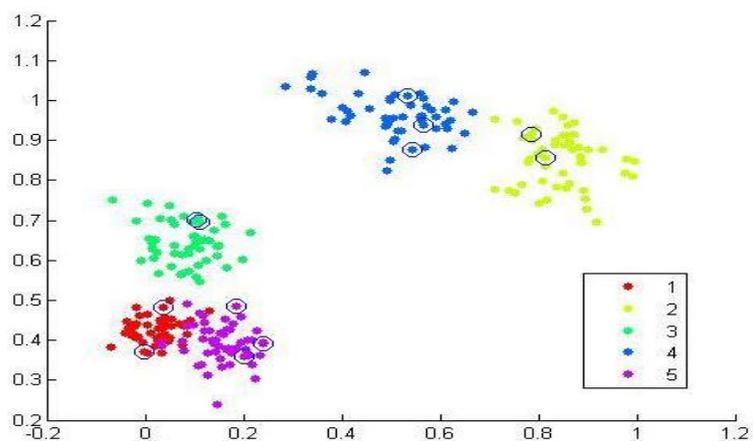


Figure 2.8:classification a multi class.

Le vecteur de support , qui apparaît dans le nom de l'algorithme du vecteur de support , est le point de données le plus proche de la ligne séparant l'ensemble de données. Ce vecteur d'appui joue un rôle important pour les SVM en décidant quelle ligne divise l'ensemble de données. Si cette ligne de division est déterminée, la classe est alors décidée en fonction du côté de la ligne où se trouvent les données que vous voulez classer. S'il y a des données à gauche de la ligne dans le graphique ci-dessus, il s'agit de la classe vert et si elle se trouve à droite, de la classe bleue.

Mécanisme du SVM :

- La **surface de séparation** est donc l'hyperplan : $w \cdot x + b = 0$
 - Elle est valide si $\forall i u_i h(x_i) \geq 0$
 - L'hyperplan est dit sous forme canonique lorsque $\min |w \cdot x + b| = 1$
- ou encore : $\forall i u_i (w \cdot x_i + b) \geq 1$

Hyperplan de plus vaste marge

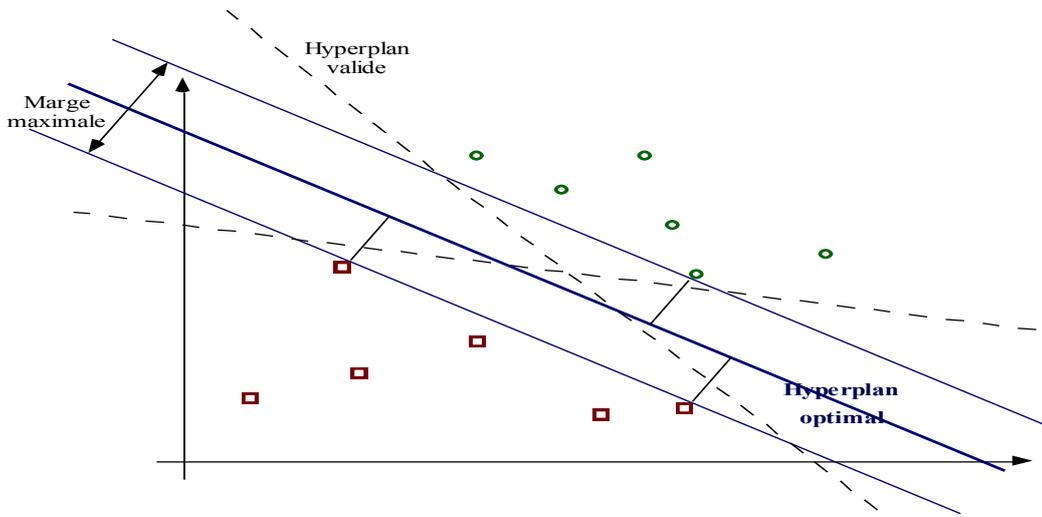


Figure 2.9: Hyperplan à marge maximale et marges pour un SVM formé avec des échantillons de deux classes. Les échantillons sur la marge sont appelés les vecteurs de support.

Optimisation de la marge

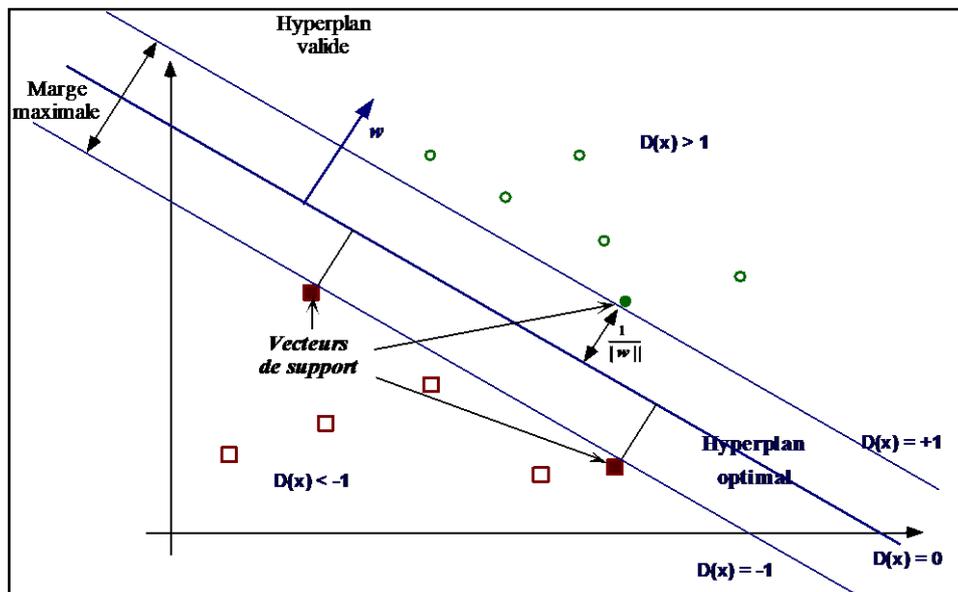


Figure 2.10: Hyperplan pour optimisation de la marge maximale et marges pour un SVM formé avec des échantillons de deux classes. Les échantillons sur la marge sont appelés les vecteurs de support.

- La distance d'un point à l'hyperplan est : $d(x) = \frac{|w \cdot x + w_0|}{\|w\|}$
- L'hyperplan optimal est celui pour lequel la distance aux points les plus proches (*marge*) est maximale. Cette distance vaut $\frac{2}{\|w\|}$
- Maximiser la marge revient donc à minimiser $\|w\|$ sous contraintes :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \forall i \quad u_i (w \cdot x_i + w_0) \geq 1 \end{cases} \quad (2.22)$$

EXPRESSION PRIMAIRE

- Il faut donc déterminer w et w_0 minimisant : $n(w) = \frac{1}{2} \|w\|^2$
- afin de maximiser le pouvoir de généralisation
- sous les contraintes (hyperplan séparateur) : $u_i [w \cdot x_i + w_0] \geq 1, i = 1, \dots, n$

EXPRESSION DUALE :

Méthode des multiplicateurs de Lagrange

$$\begin{cases} L(w, w_0, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i \{ (x_i \cdot w + w_0) u_i - 1 \} \\ \forall i \quad \alpha_i \geq 0 \end{cases} \quad (2.23)$$

Problème dual

$$\begin{cases} \max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j u_i u_j (x_i \cdot x_j) \\ \forall i \quad \alpha_i \geq 0 \\ \sum_{i=1}^l \alpha_i u_i = 0 \end{cases} \quad (2.24)$$

5.1.1.2 K-NearestNeighbor

KNN (*K-NearestNeighbor*) est un modèle qui classe les points de données en fonction des points qui lui ressemblent le plus. Il utilise des données de test pour faire une «supposition éclairée» sur ce à quoi un point non classé doit être classé.

KNN est un algorithme considéré à la fois non paramétrique et un exemple d'apprentissage paresseux. Que signifient exactement ces deux termes?

- Non paramétrique signifie qu'il ne fait aucune hypothèse. Le modèle est entièrement constitué des données qui lui sont données plutôt que de supposer que sa structure est normale.

- L'apprentissage paresseux signifie que l'algorithme ne fait aucune généralisation. Cela signifie qu'il y a peu de formation impliquée lors de l'utilisation de cette méthode. Pour cette raison, toutes les données d'entraînement sont également utilisées dans les tests lors de l'utilisation de KNN.

L'image ci-dessus montre comment KNN fonctionne lorsque vous essayez de classer un point de données en fonction d'un ensemble de données donné. Il est comparé à ses points les plus proches et classé en fonction des points auxquels il est le plus proche et le plus similaire. Ici, vous pouvez voir que le point $\{x, j\}$ sera classé comme 1 (rouge) ou 3 (vert) en fonction de sa distance de chaque groupe de points. (Machine Learning Algorithms)

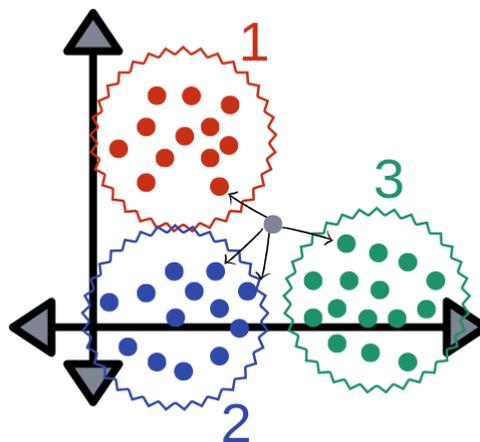


Figure 2.11: classification des données basées sur la méthode KNN

5.1.1.3 Naive Bayes

Cette méthode de classification est probabiliste. Elle consiste à utiliser les probabilités conjointes des mots et des catégories pour estimer la probabilité de classer une catégorie qui connaît. Le caractère "naïf" de cette approche est dû au fait que les mots sont considérés comme indépendants, c'est-à-dire que la probabilité conditionnelle qu'un mot connaisse une catégorie est supposée être indépendante des probabilités conditionnelles que les autres mots connaissent la même catégorie.

5.2 Apprentissage non supervisé

L'apprentissage non supervisé est un algorithme de réseau d'intelligence artificielle dont le but est d'identifier les données originales afin de comprendre la structure interne des données. Par rapport au réseau d'apprentissage supervisé, le réseau d'apprentissage non supervisé ne sait pas si le résultat de la classification est exact pendant le processus d'apprentissage, c'est-à-dire qu'il n'a pas été supervisé pour s'améliorer (lui dire quel apprentissage est le bon). Sa particularité est qu'il ne propose que des exemples d'entrée pour cette forme de réseau, et extrait automatiquement de ces exemples ses règles de catégorie possibles. Vous pouvez également l'étendre à de nouveaux cas après l'avoir étudié et vérifié.

5.2.1 Clustering

5.2.1.1 K-means

L'algorithme K-Means est un algorithme de clustering non supervisé. Il est relativement simple à implémenter.

Le regroupement K-Means est une méthode populaire pour la quantification des vecteurs N patterns en K groupes où chaque motif est une partie de la moyenne du groupe le plus proche. Cela se traduit par un partitionnement de l'espace des données en groupes.

En raison d'un ensemble de modèles d'entraînement, chaque modèle ayant des entrées à N dimensions. K-la mise en grappe des moyens vise à diviser les N patterns en K grappes afin de minimiser la somme des distances carrées au sein de la grappe. (Chris Piech, 2013)

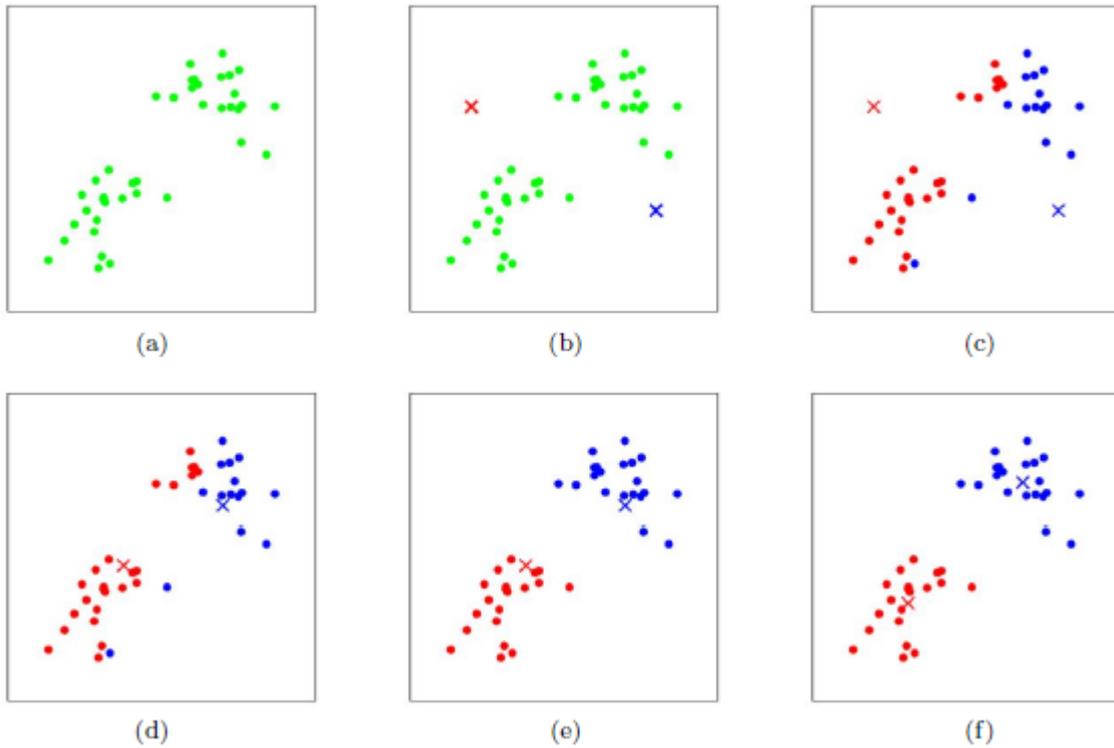


Figure 2.12 Algorithme de K-means. Les exemples d'entraînement sont représentés sous forme de points et les centres de gravité de cluster sont représentés par des croix.

Algorithm :

$it = 0$, où it = numéro d'itération et Nit = nombre total d'itérations

$it = it + 1$

Calculer le vecteur central, m_k , comme $m_k = \frac{1}{N_v(k)} \sum p: m(p) = k^x p$

où $m(p)$ est égal au numéro de cluster du p

t motif et k est égal au nombre de motifs dans k^{th}

Reclassez xps , en une seule passe de données. Si xp appartient au k

t cluster, $m(p)$ est égal à

k . $m(p)$ spécifie donc l'appartenance au cluster du p

pattern. Si seulement

les clusters changent et le numéro d'itération $<Nit$, passez à l'étape 2.

L'erreur de clustering K - Means, $E_k - means$, est

$$E_{k-means} = \frac{1}{N_v} \sum_{p=1}^{N_v} d(x_p, m_{m(p)}) \frac{1}{N_v} \sum_{k=1}^k E_k \quad (2.25)$$

5.2.1.2 Ascending hierarchical classification

Le regroupement hiérarchique (*Ascending hierarchical classification* "AHC" en anglais) est une sorte d'algorithme de regroupement qui crée un arbre de regroupement hiérarchique emboîté en calculant la similarité entre différents types de points de données.

Dans l'arbre, les points de données originaux des différentes catégories constituent le niveau le plus bas de l'arbre, et le nœud racine d'un groupe constitue le niveau supérieur de l'arbre. Il existe deux façons de créer un arbre à groupe : fusionner et diviser.

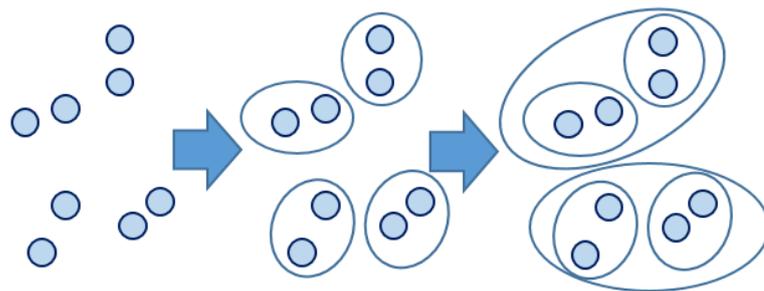


Figure 2.13: représentation graphique du clustering Hierarchical CHA

L'algorithme de clustering hiérarchique

En supposant qu'il y a n échantillons à mettre en cluster, pour l'algorithme de clustering hiérarchique, ses étapes sont:

- Étape 1: (initialisation) traiter chaque échantillon comme un cluster;
- Étape 2: Calculez la similitude entre chaque cluster;
- Étape 3: Trouvez les deux groupes les plus proches et classez-les dans une seule catégorie;
- Étape 4: Répétez les étapes deux et trois jusqu'à ce que tous les échantillons soient classés dans une seule catégorie.

6 Réduction de dimension en utilisant ACP

L'analyse en composantes principales (ACP ou PCA en anglais pour principal component analysis). La théorie de l'analyse des composantes principales est très claire. En bref, la dimension doit être choisie avec de grandes quantités d'informations et la dimension "perturbatrice" doit être omise avec moins d'informations

1 Trouver la direction de la première composante principale qui est la direction avec la plus grande variance des résultats.

2 Considérez la direction de la deuxième composante principale, c'est-à-dire la direction qui présente la deuxième plus grande variation des données, et la direction est orthogonale à la direction de la première composante principale (orthogonale si c'est un espace à deux dimensions est appelée perpendiculaire).

3 Toutes les directions des composantes principales sont déterminées de cette manière.

4 Nous pouvons obtenir les valeurs de ces composantes principales via la matrice de covariance et l'analyse des valeurs propres à l'ensemble de données.

5 Une fois que les valeurs propres de la matrice de covariance et ses propres vecteurs ont été obtenus, nous conserverons les caractéristiques N les plus importantes. Ces vecteurs de caractéristiques donnent également la structure réelle des caractéristiques les plus importantes de N, et en multipliant les données par ces vecteurs de caractéristiques, nous pouvons les convertir dans un nouvel espace.

7 Les réseaux de neurones et Deep learning

7.1 Classification des réseaux de neurones

Un réseau neuronal est une représentation des cellules neuronales (neurones) du cerveau humain et de leurs connexions. Un réseau neuronal appelé neurone artificiel par un modèle mathématique. Le réseau neuronal est composé d'une couche d'entrée, d'une couche de sortie et d'une couche cachée, et un poids "W" indiquant la force de liaison entre les neurones entre les couches est donné.

Les neurones dans le cerveau humain transmettent des informations sous forme de signaux électriques. À ce moment-là, la facilité de transmission de l'information varie en fonction de

l'intensité du lien synaptique (force de liaison). Les points fondamentaux à rappeler concernant les réseaux de neurones sont les suivants :

1 Il s'agit de structures composées de neurones divisés en plusieurs couches reliées entre elles

2 Ces systèmes peuvent généralement résoudre différents problèmes statistiques, et en particulier, en calculant les problèmes de classification à partir de l'entrée du score (ou la probabilité) de chaque classe est mis en réseau. La classe à laquelle l'objet est attribué est celle qui a la plus grande probabilité

3 Les données sont traitées et transformées à l'entrée de chaque couche en calculant une combinaison linéaire puis en appliquant une fonction non linéaire, appelée activation de fonction. Les coefficients de combinaison linéaire décrivent les paramètres (ou le poids de la couche)

4 La dernière couche de la fonction d'activation calcule les probabilités finales (classification binaire) ou la fonction softmax (classification multi-classes) ;

5 Nous attribuons également une fonction de perte à la couche de mesure finale.

6 La rétro-propagation du gradient mesure le poids des couches : calculer les paramètres en minimisant la fonction de perte progressivement contrôlée. L'optimisation est obtenue par une chute de la dernière couche à la première couche Du gradient stochastique. (Gurney, Kevin, 1997)

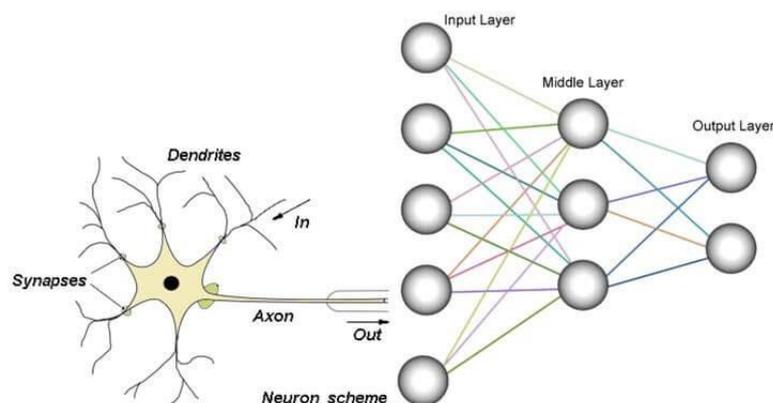


Figure 2.14: Comparaison des produits biologiques et du réseau neuronal

7.1.1 Types de réseau de neurones profond

7.1.1.1 Convolutional neural networks

Le réseau neuronal convolutionnaire (CNN) est une forme d'apprentissage profond. Ce réseau est utilisé principalement pour la reconnaissance des images. Il est appelé CNN car le poids du réseau neuronal est mesuré à l'aide de la "Convolution". La méthode de convolution est souvent utilisée dans le "filtrage spatial", qui est à la base du traitement des images.

Un CNN est constitué, comme tout réseau neuronal, d'une couche d'entrée (input layer), d'une couche de sortie (output layer) et de plusieurs couches cachées (hidden layers) entre les deux. (Stutz, David, 2014)

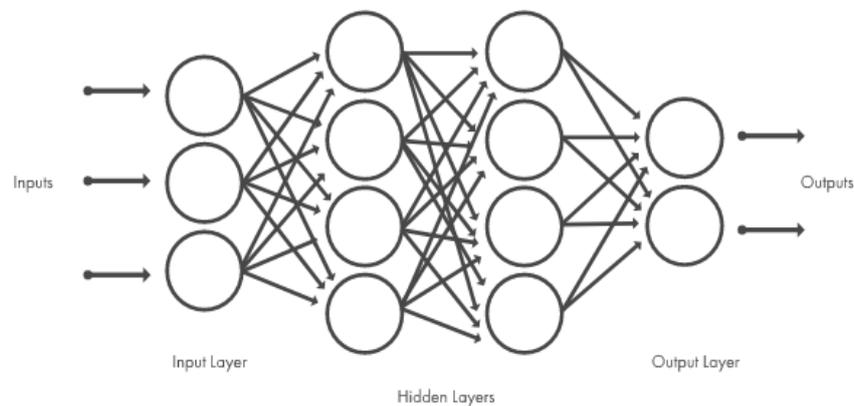


Figure 2.15: L'architecture d'un réseau neuronal, chaque couche est constituée de nœuds. Le nombre de couches masquées est facultatif.

Architecture de réseaux de neurone convolutifs

1 -Input Layer : La couche qui fournit des données sur l'entrée. Si les données d'entrée sont une image, donnez à chaque entrée 1 pixel. (Par exemple, si l'image est 32×32 [px] RGB, le nombre de dimensions d'entrée est $32 \times 32 \times 3 = 3072$)

2 -Convolutional layer : Une couche qui effectue une opération de convolution des données d'entrée et extrait les quantités de caractéristiques. La procédure de convolution est une procédure utilisée dans le "filtrage spatial", comme l'"extraction des contours" et le "traitement du flou", qui sont des exemples courants de traitement d'images. Cette procédure est également utilisée dans la couche convolutionnelle de CNN. Elle apprend automatiquement les paramètres du filtre spatial (couche convolutionnelle) en fonction de la cible que vous voulez reconnaître. En outre, CNN planifie plusieurs filtres qui doivent être

découverts à l'avance. De multiples éléments tels que le "contour" et la "couleur" peuvent être gérés en planifiant plusieurs objets, et une reconnaissance de haute précision peut être effectuée. Les données de l'image d'entrée seront augmentées par le nombre de filtres préparés. (Par exemple, s'il y a 3 filtres, le nombre d'images passera à 3)

Opération de convolution :Lorsque l'opération de convolution est effectuée sur les données d'entrée bidimensionnelles $u(x, y)$ en utilisant le filtre $h(x, y)$, la formule de calcul est la suivante.

$$u'(x, y) = \sum_i \sum_j h(i, j) u(x + i, y + j) \quad (2.26)$$

Pooling Layer :Généralement, la couche de mise en commun est utilisée après la couche de convolution.

Les données d'entrée sont comprimées dans cette couche (Down sampling).

La compression permet d'obtenir les avantages suivants et d'accroître la convivialité.

Avantages de la compression

1 Robustesse améliorée (moins vulnérable aux changements de lieu en quelques minutes)

2 Pour prévenir le sur-apprentissage.

3 Diminue le coût des calculs.

Max Pooling :La mise en pool maximale extrait uniquement la valeur maximale d'une petite zone de données d'entrée.

En compressant des informations comme celle-ci

Par exemple u' , pensez à compresser une entrée 4x4 en 2x2 avec la mise en pool Max.Les données compressées ressemblent à ceci:

$$u' = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}, u'' = \begin{bmatrix} 6 & 8 \\ 14 & 16 \end{bmatrix} \quad (2.27)$$

FullyConnected Layer :Une couche qui calcule une somme pondérée à partir des entités extraites.

La couche complètement connectée est identique au multicouche Perceptron (MLP).

Dans le cas de CNN, elle est normalement placée devant la feuille de sortie.

C'est la dernière balise.

Le nombre d'unités dans la couche de sortie doit également être en corrélation avec le nombre de groupes (le nombre à classer).

La valeur de chaque unité d'entrée (x_1, x_2) et le poids de la relation (w_1, w_2) sont fournis par le point b pour ajouter le biais. $y = f(w_1x_1 + w_2x_2 + b)$

F est la fonction d'activation.

Y est l'estimation de la valeur unitaire de la sortie.

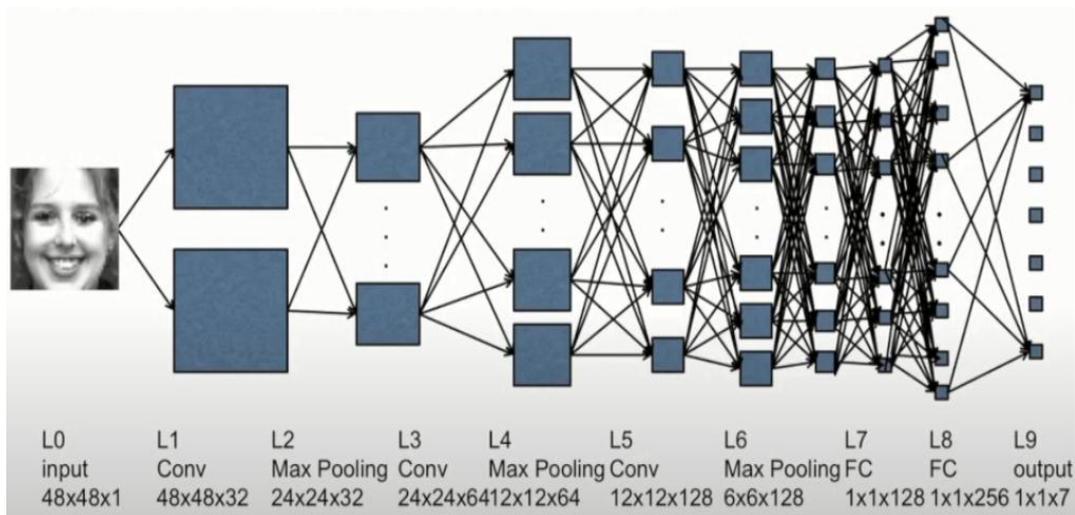


Figure 2.16: modèle pour une architecture réseau

7.1.2 Différence entre l'apprentissage automatique et l'apprentissage profond

L'apprentissage machine : Une forme d'IA qui permet aux ingénieurs et aux scientifiques de sélectionner des caractéristiques dans des modèles de données et de les former manuellement.

Apprentissage approfondi : Description des voies neuronales du cerveau humain, une sorte d'apprentissage automatique. L'algorithme apprend automatiquement les caractéristiques utiles

8 Quand utiliser Machine Learning ou Deep Learning?

Machine Learning est largement utilisé pour les projets impliquant la prévision de performances et la recherche de tendances. Dans ces cas, l'apprenant découvre des modèles avec un minimum de données qui peuvent être utilisés plus tard pour évaluer correctement les nouvelles données d'entrée. Les algorithmes les plus utilisés dans l'apprentissage machine comprennent la régression linéaire, les arbres de décision, les machines à vecteurs de support, les Bayes naïfs, l'analyse discriminante, les réseaux de neurones et les méthodes d'ensemble.(Random Forest)

Si vous avez besoin de résultats plus rapides, l'apprentissage machine sera préférable. La vitesse d'apprentissage est rapide et la puissance de traitement est limitée. Le nombre de caractéristiques et d'observations sont des facteurs importants qui déterminent le temps d'étude. Les ingénieurs qui mettent en œuvre l'apprentissage machine doivent envisager d'investir un temps considérable dans la conception et les tests des caractéristiques pour améliorer la précision du modèle.

Deep Learning est plus complexe, il est généralement utilisé dans des projets qui impliquent la reconnaissance d'images, la détection d'objets et l'amplification d'images et de signaux. Dans ce cas, les réseaux neuronaux profonds sont applicables et sont conçus pour extraire automatiquement des caractéristiques de données ordonnées dans l'espace et dans le temps, telles que des images et des signaux. Les algorithmes les plus utilisés dans l'apprentissage profond comprennent les réseaux neuronaux convolutionnels (CNN), les réseaux neuronaux récurrents (RNN) et l'apprentissage par renforcement (DQN, Deep Q-Network).

Les modèles d'apprentissage profond sont lents à comprendre. Le transfert de l'apprentissage par l'utilisation de réseaux qualifiés et de bases de données publiques réduit le temps d'apprentissage, mais peut être complexe à mettre en œuvre. En général, il faut entre une minute et quelques semaines pour apprendre un algorithme d'apprentissage profond, en fonction de votre matériel et de vos ressources informatiques. Les ingénieurs qui appliquent

l'apprentissage approfondi devraient envisager de passer beaucoup de temps à étudier des modèles et à apporter des améliorations à l'architecture des réseaux neuronaux profonds.

9 Conclusion

Ce chapitre a été réalisée sur la l'efficacité du système de Reconnaissance l'expression faciale, en rendant l'utilisation de l'extraction des caractéristiques du modèle binaire local (LBP). Il se compose principalement de trois parties, à savoir, la représentation faciale, l'extraction de caractéristiques et la classification.

III. Approche proposée pour la reconnaissance des émotions

1 1. Introduction

L'un des processus les plus importants est l'identification des expressions faciales. Ces expressions sont liées aux changements physiologiques des êtres humains qui rendent leur identification difficile.

Pour réaliser ce processus, nous avons exploité le descripteur d'extraction des caractéristiques basées sur le descripteur LBP. Ce descripteur est utilisé pour extraire les caractéristiques du visage, c'est-à-dire l'emplacement des sourcils, des yeux et de la bouche en exploitant des informations spatiales dans l'image qui est découpée en 4 régions (gauche, droite, haute, bas). Ensuite, l'identification des connaissances extraites est faite par le classificateur SVM afin de classifier les six expressions de l'être humain.

Dans ce chapitre nous allons tester le descripteur LBP sur les images de la base de données ck+, pour mettre en évidence ses performances et ses précisions de la reconnaissance des expressions faciales des individus. Ce descripteur est expliqué selon des conditions de traitement prédéfini de notre approche proposée, et également aussi en fonction du type de classificateur utilisé (SVM).

2 Reconnaissance des émotions ?

La reconnaissance des émotions est un outil utilisé dans un logiciel qui permet à un ordinateur d'utiliser une dispense d'image sophistiquée pour "examiner" les sentiments sur un visage humain. Les développeurs ont testé un amalgame de formules sophistiquées et de techniques de traitement d'images qui se sont matérialisées au cours de la dernière décennie. Les logiciels de reconnaissance d'émotions ont connu une croissance très habile avec les innovations actuelles. En outre, sa capacité à suivre les premiers regards du visage pour des sentiments tels que le bonheur, la déception, la surprise, la rage, etc. (Pelachaud, 2011)

3 Approches de la reconnaissance des émotions

Les défis de reconnaissance des émotions analysent souvent les représentations humaines multimodales telles que le texte, l'audio et la vidéo. Différents types d'émotions sont détectés par l'intégration des expressions faciales, des mouvements et gestes corporels et des informations vocales. Les approches existantes de la reconnaissance des émotions pour classer des types d'émotions spécifiques peuvent généralement être divisées en trois catégories principales : les techniques basées sur la connaissance, les méthodes statistiques et les méthodes hybrides.

1 -Techniques basées sur les connaissances : Les techniques basées sur les connaissances (parfois appelées techniques basées sur le lexique) utilisent les caractéristiques sémantiques et syntaxiques des connaissances du domaine et du langage pour détecter des types d'émotions spécifiques. Cette approche utilise généralement des ressources basées sur les connaissances dans les processus de classification des émotions tels que WordNet, SenticNet, ConceptNet et EmotiNet. L'un des avantages de cette approche est l'accessibilité et l'économie induites par la haute disponibilité de ces ressources fondées sur la connaissance. D'autre part, la limitation de cette technique est qu'elle ne peut pas gérer les nuances conceptuelles et les règles linguistiques complexes.

2- Techniques statistiques : impliquent généralement l'utilisation de différents algorithmes d'apprentissage automatique supervisé dans lesquels un grand ensemble de données annotées est entré dans l'algorithme pour que le système apprenne et prédise divers types d'émotions. Cette technique implique généralement deux ensembles de données, un ensemble d'apprentissage et un ensemble de test. Des ensembles d'apprentissage et des ensembles de test sont utilisés pour entraîner les attributs des données, ces derniers étant utilisés pour valider les performances des algorithmes d'apprentissage automatique. Les algorithmes d'apprentissage automatique fournissent généralement une précision de classification plus raisonnable que d'autres approches, mais l'un des défis pour obtenir de bons résultats dans le processus de classification est d'avoir un ensemble de formation suffisamment large. Il est nécessaire. Les algorithmes d'apprentissage automatique les plus couramment utilisés incluent SVM (Support Vector Machines), Naive Bayes et Maximum Entropy. L'apprentissage profond est également largement adopté pour la reconnaissance des émotions. Les algorithmes d'apprentissage profond bien connus incluent diverses architectures de réseaux naturels artificiels (ANN) tels que CNN (Convolutional Neural Network), LSTM (Long

Short-Term Memory) et ELM (Extreme Learning Machine). La popularité des approches d'apprentissage profond dans le domaine de la reconnaissance des émotions peut être largement due au succès des applications connexes telles que la vision par ordinateur, la reconnaissance vocale et le traitement du langage naturel.

3 - Techniques hybrides : la reconnaissance des émotions est essentiellement une combinaison de méthodes fondées sur la connaissance et de méthodes statistiques qui tirent parti des propriétés complémentaires des deux techniques. Les travaux qui appliquent un ensemble d'éléments de langage et de méthodes statistiques axés sur les connaissances comprennent l'informatique conceptuelle qui utilise tous les deux des ressources SenticNet basées sur les connaissances au niveau conceptuel. Le rôle de ces ressources basées sur les connaissances dans la mise en œuvre d'une approche hybride est très important dans le processus de classification des émotions. Les technologies hybrides ont tendance à avoir de meilleures performances de classification plutôt que d'utiliser des méthodes basées sur la connaissance ou des méthodes statistiques de manière indépendante, car elles bénéficient des avantages offerts par les approches basées sur les connaissances et les approches statistiques. Cependant, l'inconvénient de l'utilisation de la technologie hybride est la complexité de calcul du processus de classification.

3.1 Approche proposée

1.1 Extraction de caractéristiques

1.1.1 Descripteur LBP et Processus du système de la reconnaissance des émotions

- **Descripteur LBP :**

Le descripteur LBP est principalement proposé pour la première fois pour analyser de texture et du traitement de classification des données.

Dans notre mémoire, nous avons proposé une nouvelle approche de classification des émotions d'images. Généralement, une image est représentée par un descripteur de caractéristiques et une méthode d'apprentissage automatique est utilisée pour apprendre à discriminer les classes des images en fonction de ces descripteurs de caractéristiques. Le

descripteur de caractéristiques peut être extrait de manière globale à partir de l'image entière ou locale basée sur des modèles de partitionnement de l'image en plusieurs partitions.

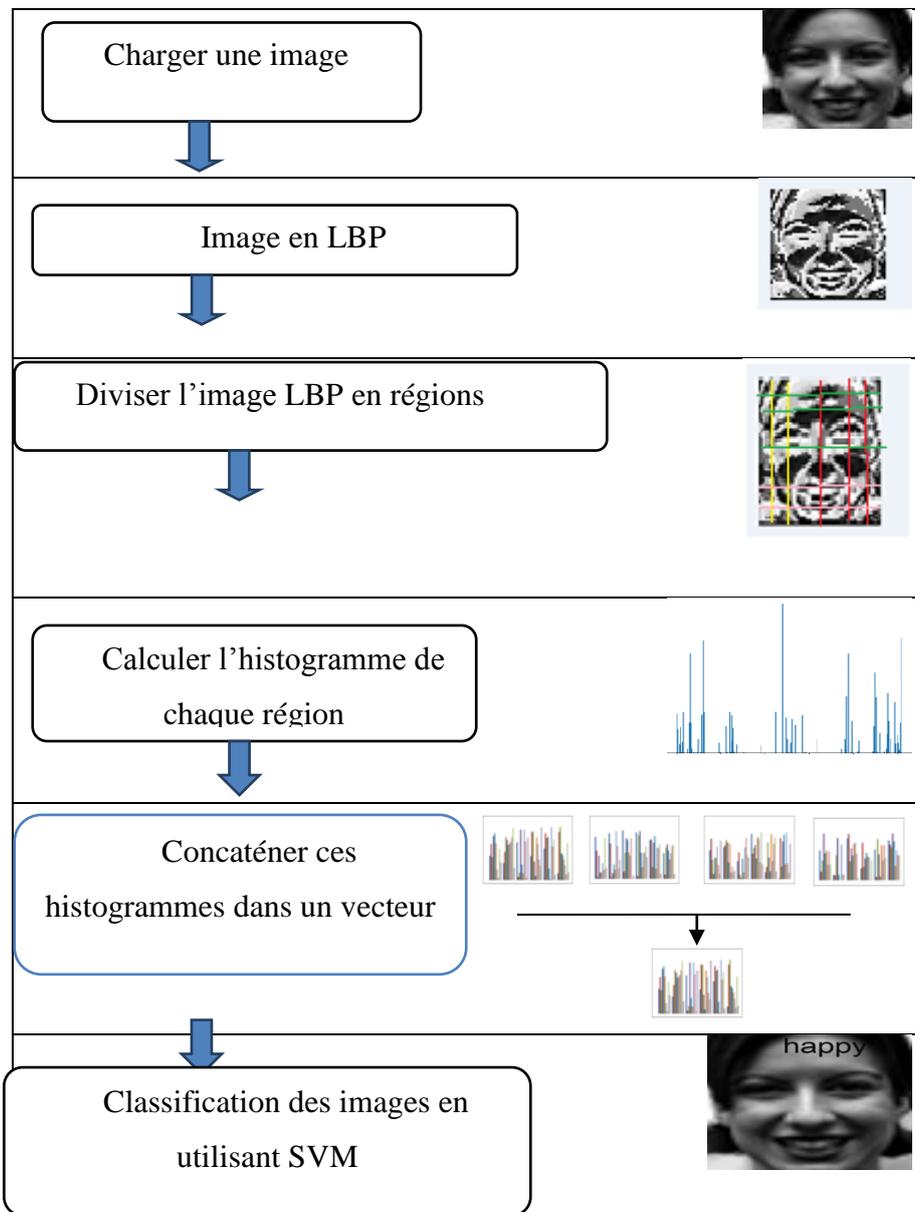


Figure 3.1: Architecture d'un système de reconnaissance des expressions faciales

L'histogramme global est l'un des moyens le plus couramment utilisé pour capturer les informations visuelles d'une image. Il présente de nombreux avantages, tels que l'invariance aux rotations de l'image et la robustesse des translations de l'image autour de l'axe de visualisation.

Les résultats de classification qui utilise l'histogramme globale n'étaient pas très prometteurs, ce qui a conduit à de nombreuses variantes des modèles de partitionnement d'une image.

- **Modèle découpage d'image**

Dans cette section, nous décrivons l'approche utilisée pour extraire les caractéristiques et capturer des informations spatiales dans une image. L'image prise est divisée de manière successive en plusieurs partitions, chaque nouvelle partition est incluse dans la partition précédente d'une façon hiérarchique. Ces partitions sont considérées comme des niveaux ($L_0, L_1, L_2, L_3, L_4, \dots$ etc) voir la figure1. La largeur et la longueur de différents niveaux sont calculés par l'équation (3.1); (3.2).

x_{L_0} est largeur de l'image I (niveau 0)

y_{L_0} est longueur de l'image I (niveau 0)

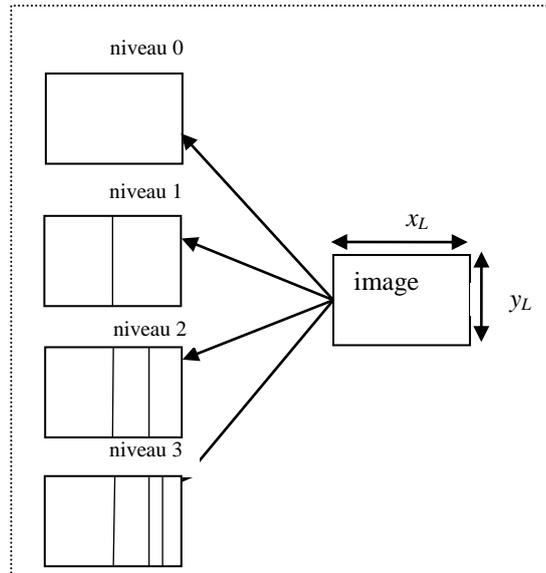


Figure 3.2: Exemple sur la construction les 4 niveaux de coté droite.

$$x_{L_k} = \frac{1}{2} x_{L_{k-1}} \quad k = 1..N \quad (3.1)$$

$$y_{L_k} = \frac{1}{2} y_{L_{k-1}} \quad k = 1..N \quad (3.2)$$

Dans notre approche, nous avons proposé quatre types de partition hiérarchiques pour calculer des caractéristiques locales qui sont basées sur les ongles d'une image. Dans le partitionnement gauche et droite en applique équation 3.1 et y_{L_k} égale y_{L_0} , et le partitionnement haut et bas en applique équation 3.2 et x_{L_k} égale x_{L_0} voir figure 2 dont

G=GAUCHE, D=DROITE, H=HAUT, B=BAS. Ensuite, des histogrammes locaux différents sont calculés pour chaque niveau.

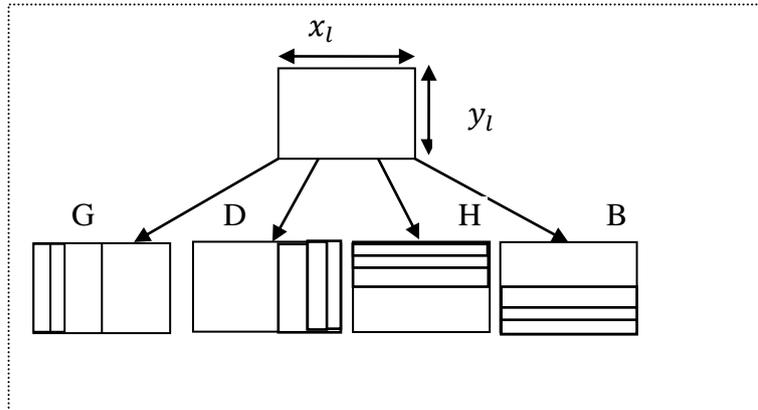


Figure 3.3: Modèle de découpage d'image

Nous pensons que cette approche enrichit la description sémantique de l'information visuelle. Parce qu'elle fournit des informations supplémentaires à l'histogramme pour capturer la distribution spatiale du contenu de l'image.

L'histogramme de caractéristiques globales est utilisé afin de décrire une image simple et permet de représenter un objet entier avec un descripteur. En revanche, les histogrammes locaux sont calculés dans plusieurs régions afin de donner plus d'importance aux informations spatiales des objets dans une image.

Avec ces avantages, l'approche de découpage d'image offre plus de pouvoir discriminant pour la reconnaissance de visage que d'autres approches.

- **Représentation l'histogramme de chaque région :**

Nous construisons une représentation en utilisant quatre niveaux hiérarchiques spatiales, voir Fig. 3.1. En général, l'histogramme global et plusieurs l'histogramme locales pour décrire des caractéristiques des objets dans une image.

Une fois les histogrammes calculés à chaque niveau dans une image, ils sont combinés pour former un seule histogramme qui représente l'image. Dans notre approche, trois niveaux de résolution spatiales ont été choisis, à savoir les niveaux sont 0, 1, 2 et 3 respectivement (voir figure 3.3).

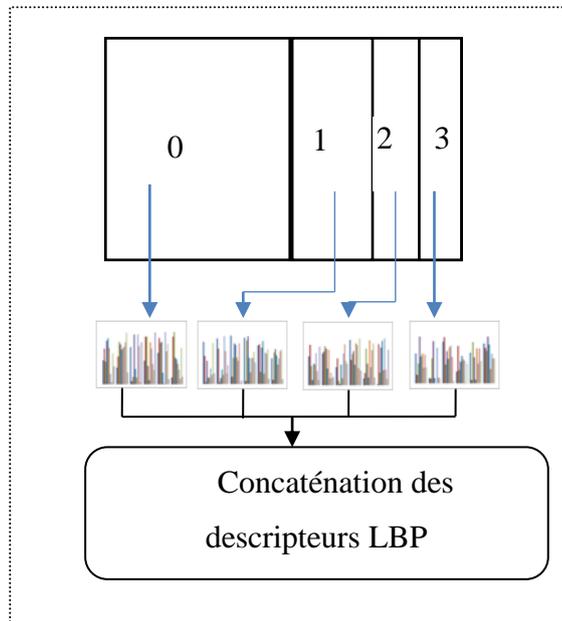


Figure 3.4: Concaténation des régions pour avoir le descripteur

L'approche de découpage d'image utilise un modèle de partitionnement successif pour construire plusieurs niveaux de résolution spatiale dans l'image.

Chaque histogramme de chaque niveau 1 est utilisé pour capturer des informations spatiales dans l'image.

En à constater que le niveau 3 avais le même résultat que le niveau 2 avec une grand taille et gaspiller plusieurs temps

Enfin, les histogrammes de l'image à tous les niveaux sont concaténés pour former un seul descripteur qui englobe l'histogramme global et local pour décrire l'image.

- **Classificateur SVM :**

Le SVM a une bonne précision de classification même s'il a utilisé n ensemble modeste de données de formation.

Le modèle de classification montré dans la figure 3.5 utilise le descripteur calculé à partir d'un seul angle de l'image. En outre, lorsque les quatre angles de l'image sont utilisés, le descripteur de caractéristiques de l'approche de découpage d'image augmentera la taille du

descripteur final qui représente l'image. Le descripteur devient très grand et le temps d'exécution devient important pour l'apprentissage et le test du SVM afin de reconnaître le visage. Cette approche peut diminuer des performances de la reconnaissance de visages.

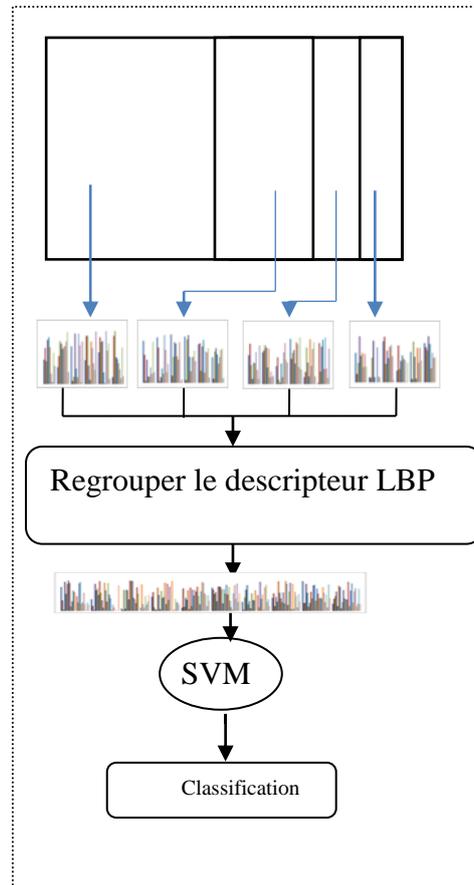


Figure 3.5: Les histogrammes sont combinés à tous les niveaux aux coté droite et une SVM est utilisée pour la classification.

Le système forme d'abord le classificateur SVM sur le descripteur calculé en utilisant le découpage de chaque partie de l'image. Dans ce cas, le classificateur SVM estime les valeurs de la classe d'un visage donnée.

- **Réduire la taille de l'histogramme en utilisant PCA :**

La combinaison des histogrammes de tous les parties dans l'image peut être plus discriminante, car elle utilise les quatre partie de découpage d'image qui capable de capturer différentes informations permettant de reconnaître le visage.

Cependant, la dimension du descripteur calculé par la partie de découpage d'image qui représente l'image complet peut devenir très importante.

Ainsi, afin de réduire la taille de ce descripteur, l'Analyse en composant principal est proposé pour ce faire. PCA est un algorithme classique de réduction de dimension largement appliqué dans le domaine du traitement visuel par ordinateur et il peut donner de très bons effets de reconnaissance de visage.

La figure 3.6 montre notre deuxième modèle de classification

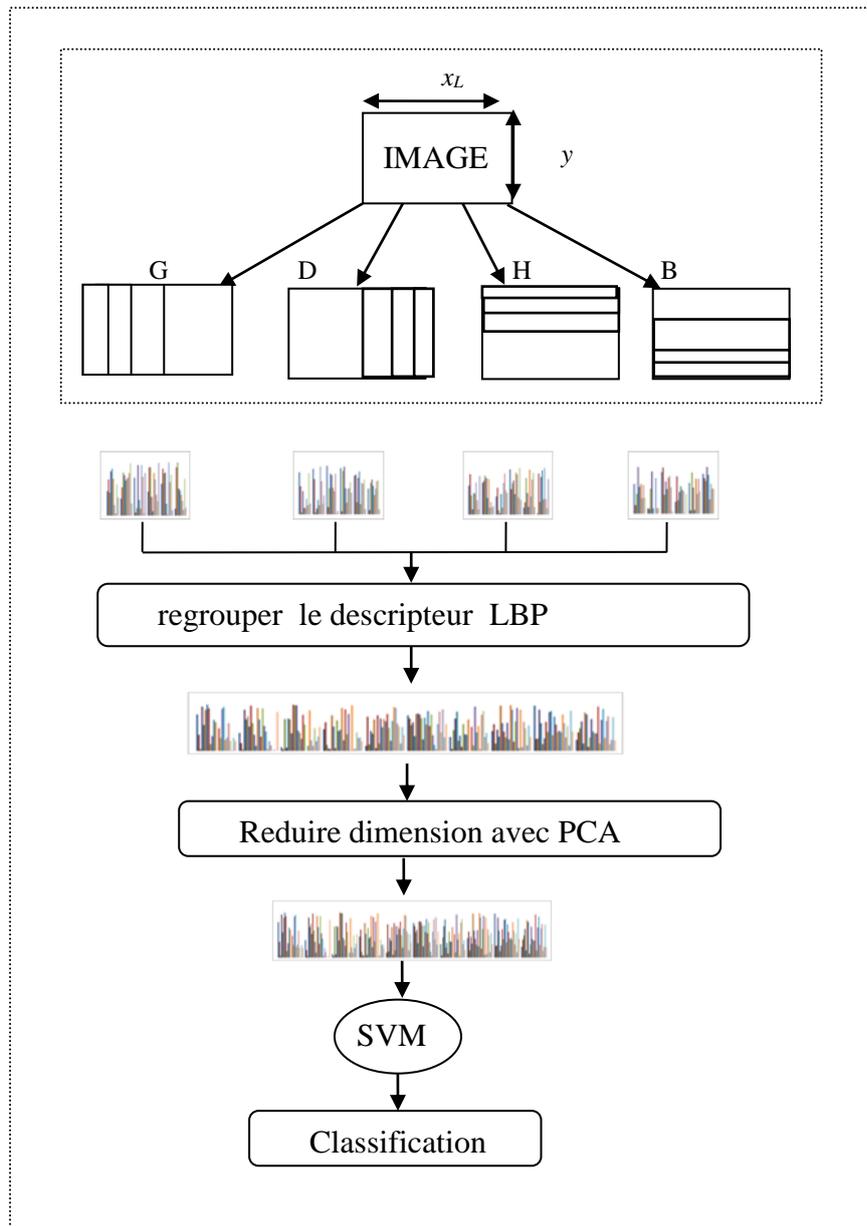


Figure 3 6: la classification image a travers PCA

3.1.1 Expériences et résultats

3.1.1.1 Basse de donnée

Dans notre travail, nous avons utilisé la base de données CK+. (kaggle) C'est une base de données d'image de visage gratuite. Elle contient sept expressions faciales posées par différents modèles féminins et masculins au total 981 images. Tous les personnes dans cette base ont les sept émotions universelles (colère, dégoût, peur, joie, tristesse, surprise et mépris), chaque expression contient 2 ou 3 échantillons de l'individu, et totalement on a 981 images de taille 48×48 pixel.

Base de données	expression faciales	Nombre de sujets	Nombre d'images	Couleur grise	Résolution,
Cohn-Kanade (CK +)	Tristesse, surprise, la joie, la peur, la colère, le mépris et le dégoût	123	981	Gris	48 * 48

Tableau 3.1: tableau représente la base de données CK+

Dans nos expériences sur la base de données CK+, nous avons utilisé 981 images (sans les images neutres) divisées comme suit :

Base d'apprentissage :

Dans notre cas, nous avons 70% échantillons, trois ou deux images par individus pour chaque émotion. Ensuite les images de différents sujets appartenant à une même classe d'émotions soient regroupées pour former la base de données d'apprentissage. Par exemple les images qui représentent l'émotion 'joie' sont situées dans une classe unique, afin que nous obtenions une base de données de sept classes représentant les sept expressions faciales.

Base de test :

Nous avons choisi une ou deux images de chaque personne pour chaque expression, donc notre base de test est composée de 30% images avec différentes expressions faciales des individus. À chaque fois on compare un visage de test avec la base d'apprentissage.

2.2.2 Résultats obtenus

Nous avons divisé les résultats de la reconnaissance des émotions en deux catégories : la première catégorie concerne la précision et le temps d'apprentissage du descripteur LBP et ces différentes régions. La deuxième catégorie concerne la réduction de la taille du descripteur en utilisant la technique PCA. Pour évaluer la performance de notre approche, nous avons utilisé l'équation 3.3.

$$Précision = \frac{T_p}{T_p + F_p} \quad (3.3)$$

TP : True Positive

FP : False Positive

1/ La première catégorie

- Niveau 0:

	LBP
temps	0.36
précision	75.2

Tableau 3.2: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 0

- Niveau 1:

L :gauche R :droite D :bas T :haut

LR :gauche et droite LRD :gauche et droite et bas

LRDT :gauche et droite et bas et haut

	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
Temps	0.40	0.40	0.32	0.36	0.48	0.59	0.79
Precision	78.3%	80%	86.7%	84%	82.7%	90.5%	93.5%

Tableau 3.3: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 1

- **Niveau 2:**

	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
temps	0.48	0.48	0.42	0.48	0.81	0.93	1.3
precision	85.7%	87.7%	95.5%	88.4%	91.1%	95.6%	96.6%

Tableau 3.4: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 2

- **Niveau 3 :**

	LBP	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
temp	0.55	0.76	0.85	0.84	0.82	1.1	1.6	2.2
precision	75.2%	84.7%	72.2%	79.6%	74.6%	90.8%	91.8%	96.6%

Tableau 3.5: Matrice de précision utilisant le descripteur LBP et le classifieur SVM sans PCA au niveau 3

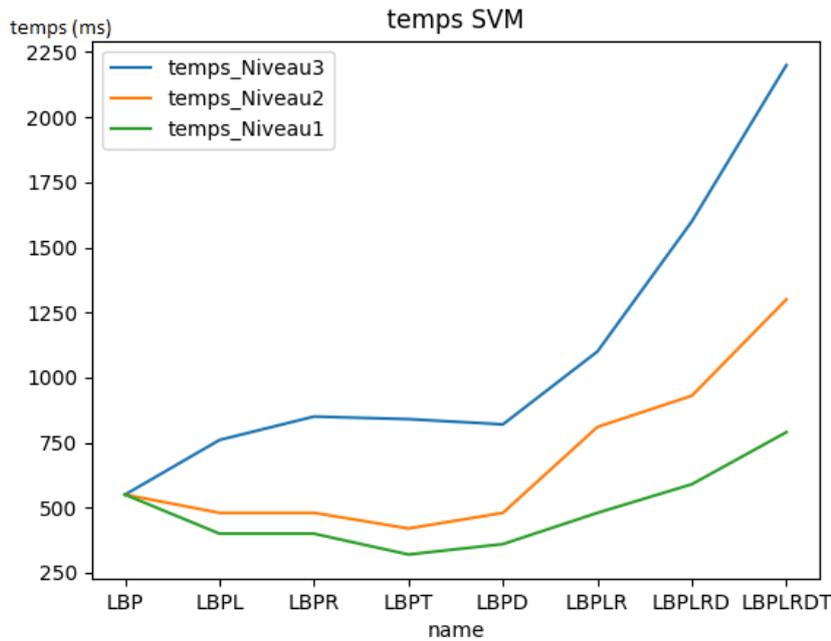


Figure 3.7: Représentation du temps de modèle apprentissage aux 3 niveaux sans PCA

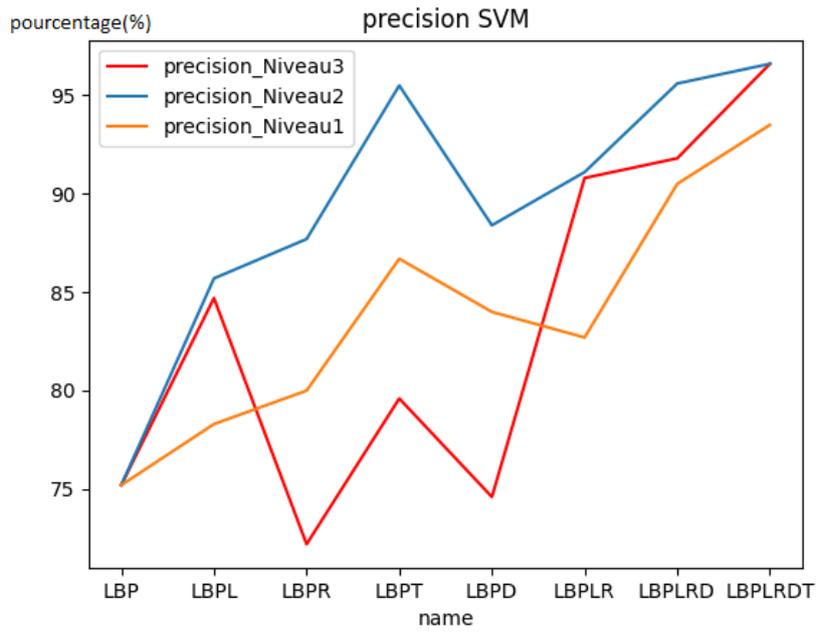


Figure 3.8: Représentation du précision de modèle apprentissage aux 3 niveaux sans PCA

2/ La deuxième catégorie

- Niveau 1:avec PCA

	LBP	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
temp	0.36	0.38	0.40	0.35	0.36	0.21	0.19	0.22
precision	75.2	78	80	84	86	82	89	92

Tableau 3.6: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 1

- Niveau 2:avec PCA

	LBP	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
temps	0.36	0.18	0.2	0.13	0.18	0.2	0.2	0.26
precision	75.2	86.6	87.7	92.5	86.4	89.4	95.2	95.9

Tableau 3.7: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 2

- Niveau 3:avec PCA

	LBP	LBPL	LBPR	LBPD	LBPT	LBPLR	LBPLRD	LBPLRDT
temps	0.4	0.22	0.32	0.29	0.33	0.25	0.26	0.34
precision	75.2	86.7	63.3	77.2	65	89.4	86.4	96

Tableau 3.8: Matrice de précision utilisant le descripteur LBP et le classifieur SVM avec PCA au niveau 3

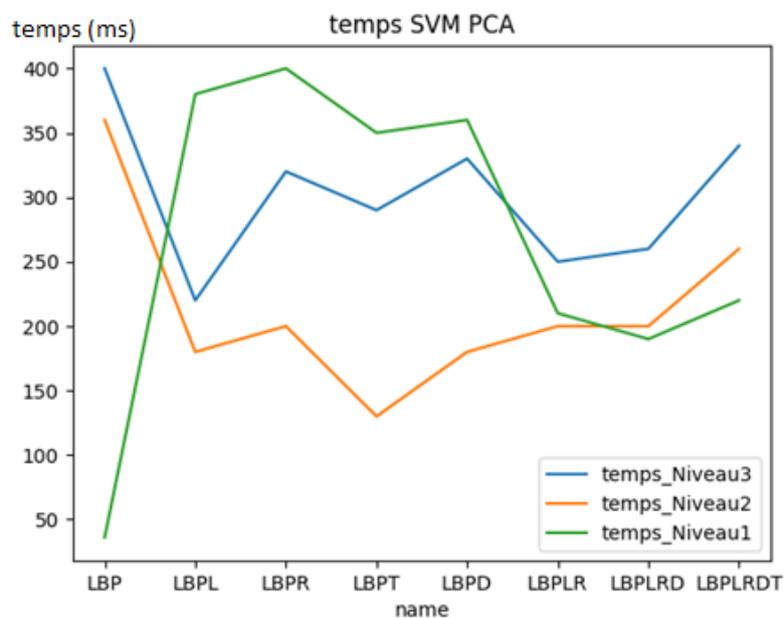


Figure 3.9: Représentation du temps de modèle apprentissage aux 3 niveaux avec PCA

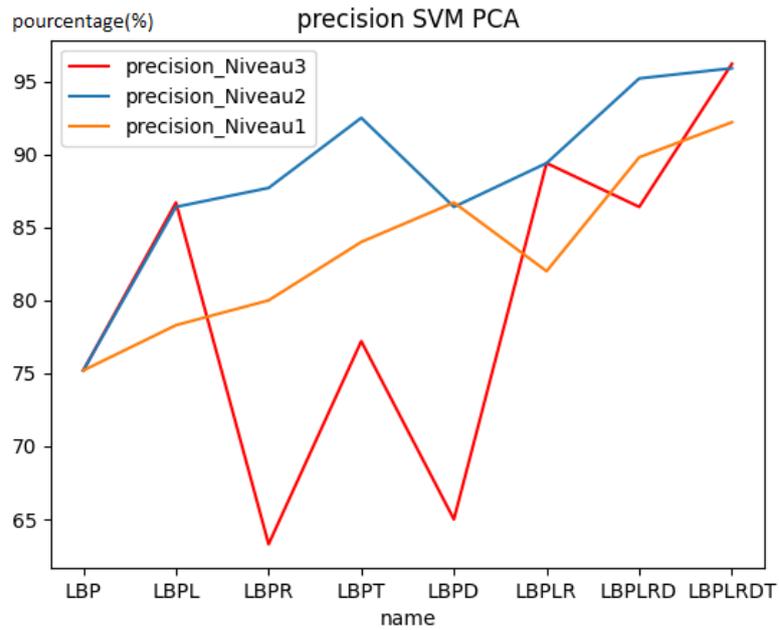


Figure 3.10: Représentation du précision de modèle apprentissage aux 3 niveaux avec PCA

Discussion de résultats :

La première partie de notre expérience est basée sur l'utilisation du descripteur LBP et ces différents dérivés hiérarchique pour extraire les données faciales. Le taux de reconnaissance de notre système qui est basé sur le descripteur LBP et ces différents dérivés donne 96.6% avec le classificateur SVM. L'utilisation de PCA sur les images de la base de données CK+ permet de garder le taux de précision et diminuer le temps de calculs.

Après avoir analysé les résultats précédents, nous avons observé que le niveau 3 donne le même résultat que le niveau 2 avec PCA. Ce dernier joue un rôle très important dans l'amélioration des performances du système de classification.

Comparaison :

Le tableau 3.9 montre les résultats de comparaison de notre méthode par rapport l'approche spatial pyramid matching qui utilisent le descripteur LBP.

SVM+PCA+SPM+LBP	Temps = 0.18	Precision=91.5%
SVM+SPM+LBP	Temps = 0.78	Precision=92.87%

Tableau 3.9: Matrice de précision utilisant le descripteur SPM et le classifieur SVM avec et sans PCA au niveau 1

Les résultats de comparaison de notre approche par rapport à l'approche de SPM montrent que les résultats obtenus sont très encourageants.

Dans une étude comparative de la reconnaissance des expressions faciale, la reconnaissance a basée sur les caractéristiques de LBP, donné un résultat de 93.5%, où la classification a été faite par le classificateur SVM, sur la base de données CK+.

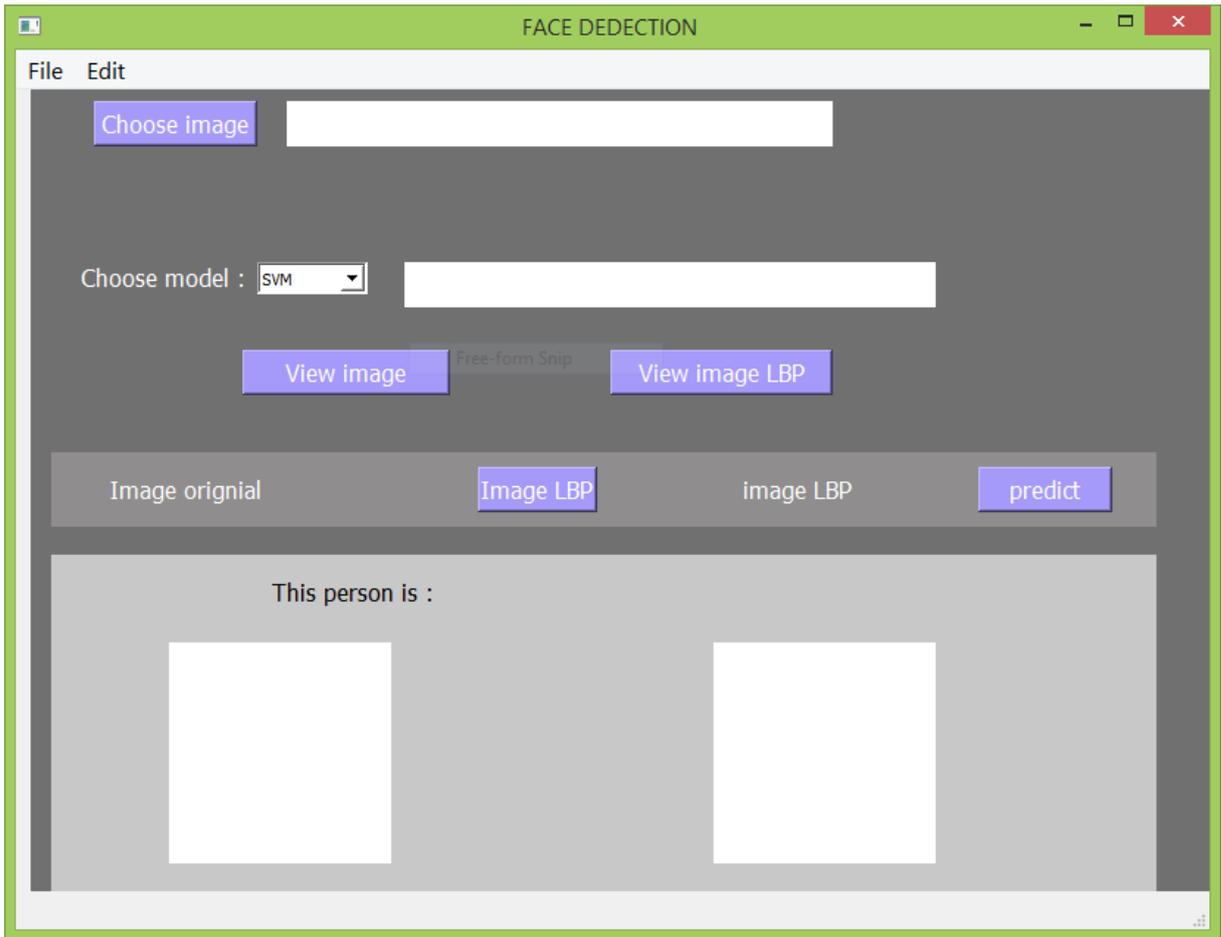
Une autre étude basée sur le SPM, où le degré de reconnaissance des expressions faciales a atteint 92,87%. Ils sont également basés sur le classificateur SVM pour la classification des expressions faciales, sur la base de données CK+.

Nous avons observé que la performance de notre système était très bonne, ceci est illustré par la comparaison des résultats dans le tableau .Où les résultats de notre travail, sont les meilleurs résultats de la reconnaissance des expressions faciales sur la base de données CK+.

Implémentation :

Dans la mise en œuvre d'un système de reconnaissance faciale, nous avons conçu une application à l'aide du langage PYTHON de la version (3.8.3). Le résultat est affiché directement dans des boîtes spécifiques selon le type de descripteur LBP. Le processus de classification est effectué par le classificateur SVM en raison de son efficacité.

L'application de reconnaissance des expressions faciales, que nous avons réalisée, a donné des bons résultats dans la reconnaissance de l'émotion des différentes images testées. La figure ci-dessous illustre la forme finale de l'application réalisée à travers notre étude avec quelque exemple de différentes expressions faciales :



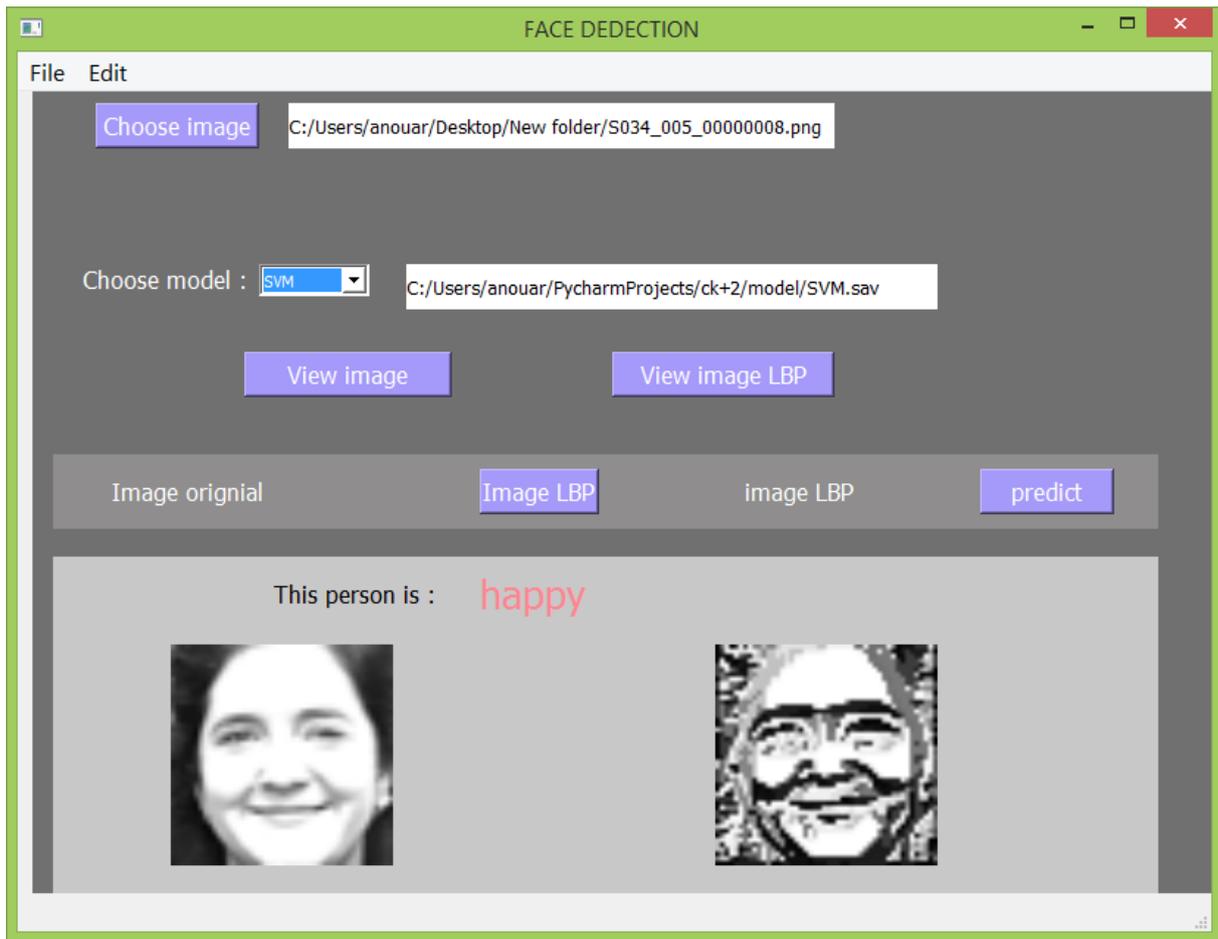


Figure 3.11: Interface de l'application de reconnaissance des expressions faciales

4. Conclusion

Dans ce chapitre, nous avons présenté les différentes étapes de l'implémentation de notre système de reconnaissance des expressions faciale

Nous avons effectué plusieurs expériences en se basant sur différentes aspect. Nous avons varié les traitements des images avec l'égalisation d'histogramme, et les méthodes d'extraction des caractéristiques faciale LBP basson sur l'approche proposée (découpage d'image), et aussi la méthode de classification multi-class qui est basée sur l'un des classificateur SVM

Les résultats obtenus pour la classification des émotions à l'aide du classificateur SVM étaient meilleurs et plus précis car elle me donne un pourcentage 96.6%

Conclusion générale

L'objectif de ce mémoire est de rechercher, concevoir et mettre en œuvre un modèle de reconnaissance des émotions. Où plusieurs disciplines ont été impliquées dans les stratégies d'identification des expressions faciales.

Dans notre travail nous avons présenté un état de l'art des différentes méthodes récentes dans ce domaine (LBP, HOG, BRISK, BRIEF, SIFT, SURF etc)

Après une étude approfondie des techniques utilisées pour identifier les expressions faciales, nous avons présenté les deux principaux étapes du système de reconnaissance de l'expression faciale, en commençant par l'étape de l'extraction des caractéristiques, et la classification, en utilisant les techniques que nous avons choisi et expliquées dans le 2ème chapitre.

Dans la phase d'implémentation nous avons effectué plusieurs expériences, en se basant sur le descripteur (LBP) en en appliquant une nouvelle approche (découpage d'image), et sur le type de classificateur (SVM) d'autre part,

Toutes ces expériences ont été menées sur la base de données ck+, comme décrit dans le 3ème chapitre. D'après les résultats obtenus et présentés dans le 3ème chapitre, les meilleurs résultats obtenus étaient basés sur le classificateur SVM, où le score de reconnaissance des expressions faciales a atteint 96.6% , pour le descripteur LBP. Aussi nous avons également obtenu un score de 92.87% avec le descripteur LBP et SPM et classificateur SVM pendant la comparaison.

Nous avons observé que le choix de type du classificateur, et le prétraitement d'images et l'étape de l'extraction des caractéristiques, jouent un rôle très important dans l'amélioration des résultats et augmentent de l'efficacité du système d'identification des expressions faciales,

Bibliographie

Alshazly, M. H. (2016). Image Features Detection, Description and Matching. *Studies in Computational Intelligence*. PP.28-30

David Matsumoto, Jeff LeRoux, Carinda Wilson-Cohn. (2000). A New Test to Measure Emotion Recognition Ability. *Journal of Nonverbal Behavior*. PP.179–209

Gurney, Kevin. (1997). An introduction to neural networks. *Computing & Information Technology* .PP.13-16

Herbert Bay, Andreas Ess, Tinne Tuytelaars. (2008). Speeded-up robust features (SURF.) *Proceedings of the 9th European conference on Computer Vision* 110(3),PP. 346–359.

Ioannis Pratikakis, Anastasia Bolovinou, Bassilios Gatos, and Stavros Perantonis. (2011). Semantics Extraction from Images. *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution - Bridging the Semantic Gap* .PP.8-10

M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly. (2016). Image Features Detection, Description. *Studies in Computational Intelligence*.P.22

M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly. (23 February 2016). Image Features Detection, Description and Matching. *Studies in Computational Intelligence*.P.24

Michael Calonder, Vincent Lepetit, Mustafa Özuysal. (2012). BRIEF: computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (34)7 .PP.778-792

NAZIR, M. J. (2018). Facial expression recognition using histogram of oriented gradients. *Cluster Computing*. PP.539-548

Ojala, T, Pietikäinen, & Harwood, D. (1996). A comparative study of texture 420 measures with classification based on featured distributions. *Pattern recognition*.*Pattern recognition*, 29 (1), PP.51–59.

Radu Horaud, Olivier Monga. (01/06/1995). *Vision par ordinateur*. Hermès - Lavoisier. PP.13-19

Stutz, David. (2014). Understanding Convolutional Neural Networks. *International Conference on*. PP.13-20

Web bibliographie

Pelachaud, C. (2011). *Les émotions dans l'interaction homme-machine*.

R. Fisher, S. Perkins, A. Walker and E. Wolfart. (2003). *Spatial Filters - Laplacian/Laplacian of Gaussian*. Récupéré sur <https://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm>

Chris Piech. (2013). *K Means*. Récupéré sur <https://stanford.edu/~cpiech/cs221/handouts/kmeans.html>

Machine Learning Algorithms. (s.d.). Récupéré sur Machine Learning is Magic! Part 2 | by Parshv Patel | Medium: <https://medium.com/@parshvpatel07/machine-learning-is-magic-part-2-690ad63e6e50>

kaggle: https://www.kaggle.com/shawon10/ckplus?fbclid=IwAR0xaRsW6VNyqa-Wf5pC5_b7b0Dn-koRuBL9PU41qVZUXq3VWyLGMFGIZQE